Folding Studies on Mutants of Chymotrypsin Inhibitor 2

by

Nadia Farida elMasry

University Chemical Laboratory,

Cambridge University

A dissertation submitted in fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Cambridge

July 4th, 1993
Gonville and Caius College
Cambridge

Preface

This dissertation is entirely my own work and is not a product of collaboration, except when stated so in the text. It has not been submitted, and is not concurrently being submitted, for any other degree, diploma or other qualification.

Nadia Farida elMasry

Apri la mente a quel ch' io ti paleso,
e fermalvi entro; chè non fa scienza,
senza lo ritenere, avere inteso.

Open thy mind to that which I unfold to thee,
and fix it there within; for to have understood
without retaining maketh not knowledge.

Dante *Paradiso* V:40-42

To Adina,

Without whose continuous efforts, support, encouragement, and advice, I would be no where near the position of having completed my first degree, let alone my PhD.

With much love and thanks, I dedicate this dissertation to her.

Acknowledgements

I would like to give thanks to Professor Alan Fersht for his generous support throughout my PhD, and to members of his group for their support and encouragement. In particular, I would like to thank Drs Mark Bycroft, Laura Itzhaki, Douglas Axe, and Jacqueline Matthews and Alan Jasanoff for their superb help in the often labourious task of proof-reading, and for giving very constructive criticism and advice on the contents of this thesis. I would like to convey very special thanks to both Drs Luis Serrano and Sophie Jackson for teaching me the importance of having faith in my own work as well as standing on my own two feet: lessons that shall put me in good stead for the future.

More recently, I have enjoyed the experience of collaborating with Dr Laura Itzhaki, Daniel Otzen and also with Alan Jasanoff who have succeeded in firing my enthusiasm for the CI2 project, even at this late stage. Alan, in particular, has given me many, many hours of encouragement, support and excellent advice, as well as having patiently listened to and answered my often seemingly endless string of "thesis-related" questions. At a time when I thought that I would only have goodbyes to look forward to, I have been lucky enough to make Alan's friendship and share with him a most happy time in Cambridge. Without him, this year would have been far more difficult and far less enjoyable than it has otherwise been.

It is often true that it may not be directly obvious to someone when they have had a part in shaping another's life, and I would like to take this opportunity to thank those people who have helped to shape or direct my life over the years. Louise Lyon and Dwayne Heinz both taught me a tremendous amount and made studying biology and chemistry a great deal of pleasure. I would hope to be able to pass on my enthusiasm for these subjects to others as well as they did to me. Beverly Walters provided me with a home, giving me the chance to continue studying at a time when I might otherwise have had to stop, as did my sister Adina some years later. Without the support of these two, it would have been very difficult to complete my education, and I cannot thank them enough. Rosemary Bittmann, Janette Jones and Helen Clark have all given me tremendous moral support, and more importantly, their friendship. I would like to thank them for being there when I have needed help, and for being a constant source of strength and happiness at all times. This is also true of the many friends I have had the great fortune to meet during my time at Caius, all of whom have contributed to a memorable four years in Cambridge.

Finally, I would like to thank Lorna, Vicky, Adina and my parents for being there when I have needed them the most, in spite of the fact that we have often lived far apart. My mother and father have encouraged me to do my best while always making me feel that my best was good enough for them. I have them to thank for my continuous optimism, and the strong sense that, even if matters are difficult at the moment, things invariably turn out for the best: a maxim which has helped me keep my sanity when research hasn't been working well!

Funding for this work has been generously provided by the Medical Research Council, with additional support from Gonville and Caius College.

Abstract

The thermodynamics and folding kinetics of mutants at the helix N-terminus and hydrophobic core of truncated Chymotrypsin Inhibitor 2 (CI2) (Met20-Gly83) have been studied. All mutants adhere to a two-state model for protein folding, and are destabilised relative to wild-type. Mutation of N-cap residue S31 to Ala or Gly destabilises CI2 by nearly 1 kcal mol^{-1}, with respect to both wild-type and the double mutant EA33EA34. Mutation of E33 or E34 to Gln, Asp and Asn progressively destabilises the protein from 0.3 - 1.1 kcal mol^{-1}. Deletion of one methyl(ene) group from the hydrophobic core of CI2 destabilises the protein on average by 1.3 kcal mol^{-1}, with a strong correlation between the environment of the mutation and its effect on stability. Finally, the helix N-terminus and hydrophobic core are partially formed in the transition state of CI2, with increased exposure to solvent compared to the native state.

Contents

Chapter 2: Materials and Methods

Chapter 3: Mutants of CI2 Fit the Two-State Model for Protein Folding

Chapter 6: Characterisation of the Transition State for CI2 Folding

Chapter 7: Conclusions and Future Work

List of Tables

List of Figures

Abbreviations

Ala (A)	alanine
Amp	ampicillin
Å	Ångstrom
Arg (R)	arginine
Asp (D)	aspartic acid
Asn (N)	asparagine
BSPI2	barley serine protease inhibitor 2
CI1	chymotrypsin inhibitor 1
CI2	chymotrypsin inhibitor 2
DEAE	diethyl amino ethyl
DNA	deoxyribonucleic acid
DSC	differential scanning microcalorimetry
DMSO	dimethylsulphoxide
E. coli	*Escherichia coli*
FPLC	fast protein liquid chromatography
GdnHCl	guanidinium chloride
Gln (Q)	glutamine
Glu (E)	glutamic acid
Gly (G)	glycine
His (H)	histidine
H-bond	hydrogen bond
Ile (I)	isoleucine
IPCR	inverse polymerase chain reaction
IPTG	isopropyl β-D-thiogalactopyranoside
Kan	kanamycin
Leu (L)	leucine
LB	Luria-Brentin medium

Lys (K)	lysine
MES	2-(N-morpholino)ethanesulphonic acid
Met (M)	methionine
M_r	molecular weight
NOE	nuclear overhauser effect
NMR	nuclear magnetic resonance
OD	optical density
PCR	polymerase chain reaction
Pro (P)	proline
RMS	root mean square
Ser (S)	serine
SDS-PAGE	sodium dodecylsulphate polyacrylamide gel electrophoresis
sAAPF*p*NA	succinyl-alanyl-alanyl-propyl-phenylalanyl-*p*-nitroanilide
2 x TY	2 x tryptone-yeast medium
Thr (T)	threonine
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
Trp (Y)	tryptophan
Val (V)	valine

Chapter 1

Introduction

1.1 The Structure and Function of Proteins

Proteins are a multifarious class of macromolecules which play an important role in a variety of biological processes. Most of the chemical reactions necessary to life are facilitated by protein catalysts, commonly known as enzymes. Many of the components of blood, including clotting agents, oxygen-binding haemoglobin, and the immunologically vital antibodies are proteins. Various proteins function as transport and storage units, control different types of growth and development, or form structural elements in the cell. Proteins are also involved in the regulation of gene expression, and ultimately in the production of other proteins.

The diversity of functions in which proteins are involved belies the fact that they form part of what is essentially a homogeneous group of molecules. The basic building blocks of all proteins are the twenty amino acids. These unite to form unbranched chains of up to several hundred amino acids in length, conferring on each protein a specific primary sequence which has been determined biologically by the gene encoding for that protein. This polypeptide chain then folds into a well-defined three dimensional structure which is critical to the protein's *modus operandi*. As early as 1961, it was shown that some proteins can denature reversibly *in vitro*; i.e., on going from denaturing to native conditions these proteins are able to spontaneously refold and recover their original fully functional form (Anfinsen, 1961, 1973). This has been shown to be the case for many small globular proteins, suggesting that all of the information necessary to fold from a disordered polypeptide chain to the most

thermodynamically stable native structure is present in the primary amino acid sequence. Exactly how this folding process occurs remains to be solved.

Since the deciphering of the genetic code and the advent of genetic engineering techniques, the primary amino acid sequences of over 15,000 proteins have been determined. Unfortunately, the elucidation of three-dimensional structures with the use of X-ray crystallography and NMR has proceeded at a much slower rate, as only several hundred such structures are available by comparison. The cracking of the "second genetic code," i.e., the ability to predict a protein's three-dimensional structure *ab initio*, would be invaluable in the study of proteins and hence facilitate the rational design of novel proteins. Attainment of this elusive goal, however, seems unlikely in the near future.

1.2 The Process of Protein Folding

(i) Protein Folding *in vitro*:

Remarkably little is known about the processes involved in protein folding. One approach to the unravelling of individual steps along the folding pathway has been to analyse unfolding and refolding reactions *in vitro* . Folding studies *in vitro* have shown that the reaction is usually rapid (completed within seconds or less), and for larger proteins folding begins even before synthesis is completed. Several models have been proposed (see section 1.3) to explain the spontaneous folding of many small proteins. Recent experimental evidence supports the idea that there are preferred pathways for protein folding, in which the reaction passes through discrete and highly transient intermediates. Such a process would appear to be cooperative, whereby formation of one region of secondary structure enhances the formation of the next region, and so on. The existence of parallel pathways as an alternative mechanism, however, should not be ruled out.

Analysis of the folding process of some proteins can be complicated by the presence of X-proline peptide bonds in the *cis*-conformation and by disulphide bridges. For proteins that contain X-Pro peptide bonds in the *cis*-conformation, *cis-trans*

isomerisation about these bonds can dominate the refolding process (the *trans*-conformation being somewhat favoured in the unfolded form of the protein). The slow phases observed in the refolding kinetics of proteins are frequently the result of such slow *cis-trans* isomerisation (Kim & Baldwin, 1982; Kiefhaber *et al.*, 1990). Likewise, the formation or rearrangement of disulphide cross-links is often a rate-limiting step in the folding of proteins that contain these bonds (Creighton, 1978, 1988). The presence of disulphide bonds can cause other problems as well. In lysozyme (Broadhurst *et al.*, 1991) and staphylococcal nuclease (Shortle & Meeker, 1989), intact disulphide bonds in the denatured state can result in the presence of residual structure which complicates analysis of the folding pathway because a reference state with minimal structure is usually required.

Some proteins are unable to fold without the presence of a leader peptide that is not present in the final native state of the protein. Extracellular serine proteases are synthesised as pre-pro-proteins, where the pre-sequence acts as a signal for translocation of the protein across the membrane, and the pro-sequence is thought to be involved in the correct folding of the protease. Deletion of the pro-sequence of subtilisin E, for example, results in mature but inactive enzyme (Ikemura *et al.*, 1987). Activity is restored upon exogenous addition of pro-sequence, however (Zhu *et al.*, 1989). Unfolded or partially folded proteins are not very soluble in aqueous solution, which can often lead to the premature aggregation of proteins *in vitro*. This problem can sometimes be avoided by using dilute protein solutions in experiments. Due to the high concentration of proteins in the cell, it is reasonable to assume that aggregation would be a great problem *in vivo*. It would seem, however, that aggregation in the cell is prevented with the assistance of certain proteins, to be discussed below.

(ii) Protein Folding *in vivo*:

Some proteins have been found to require the assistance of auxiliary factors, such as the enzymes protein disulphide isomerase (PDI) and peptidyl-prolyl-isomerase (PPI), in the folding of their nascent polypepetides in the cell. PDI is thought to

accelerate the formation of correct disulphide bonds during refolding (Freedman, 1984, 1989) and PPI is believed to catalyse X-Pro peptide bond isomerisation (Lang *et al.*, 1987). In addition, a class of proteins known as "molecular chaperones" appear to be essential for the correct folding and assembly of certain other proteins in the cell. Most chaperones belong to the family of heat-shock proteins (notably the hsp60 and hsp70 class), have a M_r of around 60 k per monomer and share strong sequence homology. Molecular chaperones are thought to recognise misfolded or aggregated molecules, perhaps binding transiently to partially folded polypeptide chains in order to prevent aggregation before the completion of folding (Ellis & Hemmingsen, 1989; Schmid, 1991). In *E. coli*, the chaperones GroEL and GroES are responsible for the folding of monomeric proteins and aid in the assembly of oligomers (Goloubinoff *et al.*, 1989), although the molecular mechanism is not known.

1.3 Models to Describe the Pathway of Protein Folding

(i) The Possibility of Single or Multiple Pathways:

Levinthal (1968) proposed that an inordinate length of time would be required for a protein to find its native folded state by randomly searching among all possible conformations. In an illustration of "Levinthal's paradox" it is estimated that a protein of 101 amino acids would need approximately 10^{27} years to fold into its native state by such a random search (Zwanzig *et al.*, 1992). As it takes many small globular proteins less than one second to fold, protein folding by such a random search mechanism seems implausible. It is now widely thought that there are preferred pathways of protein folding, which lead to the native conformation via specific intermediates (Kim & Baldwin, 1982; Dobson & Evans, 1988). Several models have been suggested as to the nature of such folding pathways, though the differences between these models often appear to be a matter of semantics. The models seem to fall into two broad categories: those that predict the formation of local structure before global structure and those that predict the converse.

In the "framework" model, supported by Baldwin and colleagues (Kim & Baldwin, 1982; Udgaonkar & Baldwin, 1988), the formation of stable secondary structure is an early folding event and provides the necessary framework for the subsequent formation of the complete tertiary structure. According to the framework model, folding of the protein is directed along one or relatively few well-defined sequential pathways along which unique structural intermediates may be found. The diffusion-collision or subdomain model proposes that the earliest events in folding are the formation of native-like structural domains which then collide into one another to form the final, tertiary structure. In the subdomain/collision models, if there are several initiation sites then it is possible that different, parallel pathways exist along which the protein can fold (Karplus & Weaver, 1976; Oas & Kim, 1988; Staley & Kim, 1990; Moult & Unger, 1991).

Chan & Dill (1990) support a model of protein folding in which collapse of the hydrophobic core is the primary event, followed by secondary structure formation in a series of local reorganisations. The intermediate structure might be the molten globule state, a compact intermediate which possesses native-like secondary structure with vacillating tertiary interactions (Ptitsyn, 1987; Kuwajima, 1989). Ptitsyn and Kuwajima support the idea that a "molten globule" intermediate exists on the folding pathway of all proteins, its conversion to the native state being rate-determining (Ptitsyn, 1987; Kuwajima, 1989).

Finally, the "jigsaw puzzle" model, proposed by Harrison & Durbin (1985), holds that many pathways exist in parallel, all of which are equally accessible to the unfolded protein. Analogous to the problem of assembling a single-colour jigsaw puzzle, there is no defined starting point for the folding reaction; each folding attempt follows a different path and does not converge with others to any observable structural intermediates. The jigsaw puzzle model rules out the existence of specific rate-limiting transition states along the folding pathway and tertiary structure can form without prior formation of secondary structure. In support of this model, it is possible to assemble the native structure of ribonuclease A in several different ways by association of

compact structural units which are contiguous in the primary sequence (Lesk & Rose, 1981).

(ii) Experimental Evidence for the Existence of Well-Defined Pathways:

For such models to be tested, it is necessary to examine the interactions that are made along the folding pathway, determining when and to what extent these interactions are formed. Intermediates on the folding pathway may be detected using kinetics. Kinetic intermediates are transient in nature, however, making it difficult to characterise the structure of such species. One approach to this problem has been to trap intermediates that form along the folding pathway. In a classic experiment, the presence of an early folding intermediate in ribonuclease A was demonstrated by Udgaonkar & Baldwin (1988). Using quenched-flow H/D-exchange monitored by NMR, it was possible to see at what point along the folding reaction different labile protons became protected from exchange, presumably as a result of structure formation. A similar experimental approach was used by Roder *et al.* (1988) to characterise the structural intermediates of cytochrome *c* and Bycroft *et al.* (1990) have employed this method to obtain information about the folding pathway of the small ribonuclease from *Bacillus amyloliquefaciens*, barnase. The results of the H/D-exchange experiments suggest that the formation of some secondary structure is an early folding event, preceding the formation of the final, native structure. These results support both the framework and subdomain/collision models for protein folding, although they do not rule out the possibility that the formation of secondary structure coincides with the collapse of the hydrophobic core. Random mechanisms, such as the "jigsaw puzzle model" of Harrison & Durbin, appear to be excluded. The limitations of the H/D-exchange/NMR technique are summarised by Creighton (1990, 1991). Only those amide protons protected in the final, native state of the protein can be used as folding probes since partners that provide protection during transient intermediate states are not known and, hence, the sequence of events cannot be determined directly.

Studies on fragments of the proteins ribonuclease A (Brown & Klee, 1971), BPTI (Oas & Kim, 1988) and more recently, barnase (Sancho *et al.*, 1992a, 1992b) have shown that these fragments possess some secondary structure under certain conditions. The work on fragments corroborates both the framework and subdomain models for protein folding. Kinetic studies have also been used to detect a molten globule-like intermediate along the folding pathway of several proteins, e.g., α-lactalbumin monitored by circular dichroism (Kuwajima *et al.*, 1985) and bovine carbonic anhydrase, bovine β-lactoglobulin, staphylococcus aureus β-lactamase, and human α-lactalbumin, all monitored by ANS fluorescence (Ptitsyn *et al.*, 1990).

Matouschek, Fersht and coworkers (Matouschek *et al.*, 1989, 1990, 1992a, 1992b, Fersht *et al.*, 1992) have used protein engineering techniques in combination with both fluorescence-monitored stopped-flow and H/D-exchange quenched-flow kinetics to analyse the interactions that are formed or broken in the intermediate and subsequent transition state for barnase folding. Individual mutations are positioned throughout the protein to act as probes for the formation of structure in the vicinity of the mutation. This strategy, used to characterise the intermediate and transition states and analyse the folding pathway of barnase, may be applied to other proteins.

More needs to be known about the factors which contribute to the stability of the native structure of proteins, as well as an idea of the pathway of protein folding *in vitro*, before it becomes possible to understand the molecular mechanism of such folding *in vivo*. In order for interpretation of results to be as clear-cut as possible, the systems used for these studies should be very simple. An ideal starting point for investigating the rules of protein folding are small, monomeric proteins which lack both disulphide bonds and X-proline peptide bonds in the *cis*-conformation. Small, single domain proteins often adhere to a simple two-state model for unfolding, lacking significantly populated intermediate states. The serine protease inhibitor, chymotrypsin inhibitor 2, fulfils these criteria and offers an excellent model for the study of protein folding.

1.4 CI2: A Model System for Studying the Protein Folding Problem

(i) Chymotrypsin Inhibitor 2:

Chymotrypsin Inhibitor 2 (CI2) is a member of the potato I family of serine protease inhibitors (serpins), and is isolated from the albumin fraction of seeds from the *Hiproly* strain of barley (Jonassen, 1980). The *Hiproly* strain, originally discovered as a spontaneous mutant in the world barley bank (Munck *et al.*, 1970), owes its high lysine content in part to the high level of CI2 found in the grain's endosperm. CI2 is also known as barley serine protease inhibitor 2 (BSPI-2), and shares sequence homology with eglin c derived from leeches (Seemuller *et al.*, 1977) and CI1, another member of the potato I family which is less thermally stable than CI2 and does not inhibit subtilisin as effectively (Boisen *et al.*, 1981; Svedsen, 1982; McPhalen & James, 1988).

CI2 was first observed to inhibit chymotrypsin *in vitro*, although it is in fact a better inhibitor of subtilisin BPN'. Its natural target protease *in vivo* is not yet known. Plant serpins are thought to play a defensive role, being released when the plant needs protection against wounding or attack (Ryan, 1978). It is proposed that when pests ingest these serpins, the digestive enzymes of the pest are inhibited which causes its starvation and death. Serpins act as competitive inhibitors by tightly binding to the active site of the target protease (K_i is typically 10^{-8} - 10^{-16} M) with subsequent hydrolysis of the reactive site bond of the inhibitor. The hydrolysis and release of cleaved product proceeds very slowly, however (Laskowski & Kato, 1980). A high resolution crystal structure of CI2 in complex with subtilisin (McPhalen *et al.*, 1985) supports the kinetic studies of Longstaff *et al.* (1990) which suggest that the inhibited complex represents a "frozen" Michaelis complex where the inhibitor is bound with minimal change in its structure and hydrolysis occurs slowly, if at all.

CI2 is a rare example of a protein that undergoes a reversible, cooperative, two-state transition, both thermodynamically and kinetically (Jackson & Fersht, 1991a, 1991b). Of great importance is the fact that CI2 has been successfully cloned and over-expressed in *Escherichia coli*, making the production and analysis of wild-type and

mutant CI2 realistic (Longstaff *et al.*, 1990). It lacks both disulphide bonds and X-Pro peptide bonds in the *cis*-configuration, which often complicate the analysis of protein folding. Both high resolution crystal (McPhalen *et al.*, 1985; McPhalen & James, 1987) and NMR solution structures (Kjaer *et al.*, 1987, Kjaer & Poulsen, 1987; Clore *et al.*, 1987a, 1987b; Ludvigsen *et al.*, 1992) are available for CI2. Extensive studies carried out on the folding of wild-type CI2 (Jackson & Fersht, 1991a, 1991b) have layed the groundwork for further studies using engineered mutants.

(ii) Structure of CI2:

CI2 consists of a single polypeptide chain of 83 amino acids ($M_r = 9250$) and contains no cysteine residues (Svedsen *et al.*, 1980). The primary amino acid sequence is given in Fig. 1.1 (McPhalen *et al.*, 1985). Although it is a small protein, CI2 still contains significant secondary structural elements (Fig. 1.2), including a four-stranded mixed parallel and anti-parallel β-sheet which packs against the α-helix to form the hydrophobic core. Between strands 2 and 3 of the β-sheet is the wide reactive site loop in an extended conformation. The reactive site loop is the region containing the reactive site bond (between residues Met59 and Glu60) which binds into the active site cleft of the protease that is being inhibited. CI2 is a wedge-shaped disk of approximate dimensions 28x27x19 Å, the reactive site loop forming the narrow end of the wedge (McPhalen & James, 1987).

CI2 is anomalous in the class of serpins in that it lacks the disulphide bonds which normally flank the reactive site bond and stabilise the reactive site loop of inhibitors from other families. CI2 does appear to have comparable non covalent interactions, though, involving two highly conserved arginine residues. Arg65 and Arg67 probably stabilise the conformation of the reactive site loop through hydrogen bonding and electrostatic interactions with Thr58 and Glu60 (Jandu *et al.*, 1990). Four of the five X-Pro peptide bonds in CI2 are well-defined in the crystal and NMR structures (see sections 1.5(i) - (ii)) and are in a *trans*-conformation. The fifth proline is at the N-terminus of the protein which is not defined in either crystal or NMR

structures. A single, buried tryptophan residue (Trp24) in the hydrophobic core of CI2 enables changes in the folding state of the protein to be monitored by changes in emission fluorescence. It has been shown previously using 1-D ^{1}H-NMR (Jackson & Fersht, 1991a) that the unfolding transition monitored fluorimetrically is due to the global denaturation of the protein, and not due to some local unfolding event in the vicinity of Trp24.

```
                    5                   10                  15
(GLX,VAL, SER,SER) LYS-LYS-PRO-GLU-GLY-VAL-ASN···THR GLY- ALA- GLY···ASP

                20                  25                  30
ARG···HIS-ASN-LEU-LYS-THR-GLU-TRP-PRO-GLU-LEU-VAL-GLY-LYS-SER-VAL-GLU
                  β   β   β/T   T   T   T   T   T   T   α   α   α

    35                  40                  45                  50
GLU-ALA-LYS-LYS-VAL-ILE-LEU-GLN-ASP-LYS-PRO-GLU-ALA-GLN-ILE-ILE-VAL-LEU
  α   α   α   α   α   α   α   α   α/T   T   T   T   T   β   β   β   β   β

            55                  60                  65
PRO-VAL-GLY-THR-ILE-VAL-THR-MET*GLU-TYR-ARG-ILE-ASP-ARG-VAL-ARG-LEU
  β   β                             T   T   T/β   T/β   β   β   β

    70                  75                  80              83
PHE-VAL-ASP-LYS-LEU-ASP-ASN-ILE-ALA-GLU-VAL-PRO-ARG-VAL-GLY
  β   β   T   T   T   T   β   β   β   β   β   β   β   β   β
```

Secondary Structure	Residue No.
β-sheet strand 1	22 - 24
Type III reverse turn	24 - 27
Type II reverse turn	27 - 30
α-helix	31 - 43
Type I reverse turn	43 - 46
β-sheet strand 2	47 - 53
Reactive site loop	47 - 61
Type 1 reverse turn	62 - 65
β-sheet strand 3	64 - 70
Type 1 reverse turn	71 - 74
β-sheet strand 4	75 - 83

Fig. 1.1: Primary Amino Acid Sequence of CI2 and Secondary Structure Elements; (···) indicate known sites of proteolytic cleavage, (*) reactive site, (α) α-helix, (β) β-sheet, and (T) turn.

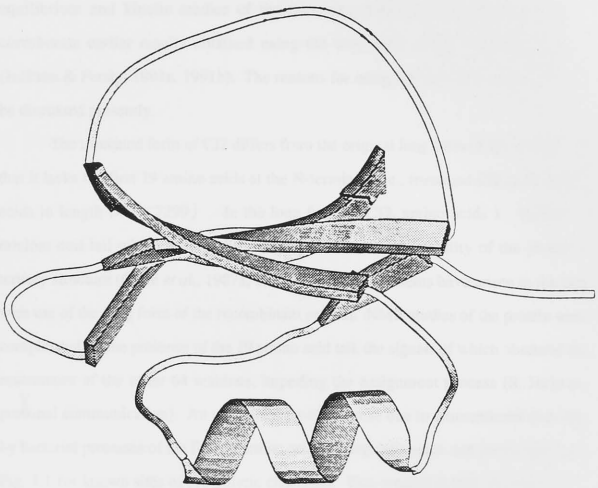


Fig. 1.2: The Backbone Structure of Chymotrypsin Inhibitor 2

(iii) Truncated Form of Recombinant CI2:

Original studies on CI2 performed in this laboratory involved the recombinant long form of the protein (83 residues). A truncated form of the recombinant protein has been used, however, for all studies on CI2 presented in this thesis. Results of equilibrium and kinetic studies of the truncated wild-type CI2 (see Chapter 3) corroborate earlier results obtained using the long form of the wild-type protein (Jackson & Fersht, 1991a, 1991b). The reasons for using a truncated form of CI2 will be discussed presently.

The truncated form of CI2 differs from the original long form of the inhibitor in that it lacks the first 19 amino acids at the N-terminus, i.e., truncated CI2 is 64 amino acids in length (M_r = 7299) . In the long form of CI2, amino acids 1 - 19 form a random coil tail which is thought not to contribute to the stability of the protein's tertiary structure (Clore *et al.*, 1987a, 1987b). Several problems have arisen in the past with use of the long form of the recombinant protein. NMR studies of the protein were complicated by the presence of the 19 amino acid tail, the signals of which obscured the resonances of the other 64 residues, impeding the assignment process (S. Jackson, personal communication). An additional problem arose due to unintentional cleavage by bacterial proteases of the first 18 amino acids during expression and purification (see Fig. 1.1 for known sites of proteolytic cleavage). This created at least three isoforms of CI2, drastically reducing the yields of a single isoform necessary for analysis.

To overcome these difficulties, the plasmid pCI2 was constructed to express a truncated form of the protein, Met20 - Gly83. NMR (Clore *et al.*, 1987a, 1987b; Ludvigsen *et al.*, 1992) and crystallographic studies of the recombinant truncated CI2 (Harpaz *et al.*, manuscript in preparation) suggest that removal of this tail does not affect the overall structure of the protein. In addition, all of the four X-Pro peptide bonds present in the truncated form of CI2 are in the *trans*-conformation.

1.5 NMR and Crystallographic Studies of CI2

(i) NMR Studies on CI2 (long and truncated form):

^{1}H NMR solution structures have been obtained for the long form of CI2 (Kjaer *et al.*, 1987; Kjaer & Poulsen, 1987) as well as for a proteolytically cleaved form of the protein (residues 19 - 83; Clore *et al.*, 1987a, 1987b). Most recently, an NMR solution structure of the recombinant truncated form of CI2 (residues 19 - 64) has become available (Ludvigsen *et al.*, 1992). These results suggest that the overall structure of the protein has not changed with removal of the 19 amino acid N-terminal tail.

(ii) Crystal Structures of Mutant and Wild-type CI2:

CI2 was the first member of the potato inhibitor I family of serine proteinase inhibitors to have its structure elucidated by X-ray crystallography. The structure of long-form wild-type has been solved to high resolution both for CI2 as free inhibitor (resolved to 2.0 Å by McPhalen & James, 1987) and in complex with subtilisin Novo (resolved to 2.1 Å by McPhalen *et al.*, 1985). More recently, the crystal structure of the truncated form of wild-type CI2 has been solved to 2.0 Å resolution (Y. Harpaz, unpublished data) and corroborates evidence that the structure of the truncated form does not differ significantly from that of the long form. Intermolecular crystal contacts involving residues near the helix N-terminus of this structure, however, make difficult its use when interpreting the effect of helix mutants on structure and stability of CI2.

Structures have also been obtained for several mutants of the truncated form of CI2. These include a 2.2 Å resolution structure of the hydrophobic core mutant IV76 (Jackson *et al.*, 1993a) and helix mutants SG31 (2.0 Å resolution structure), EA33EA34, SG31EA33EA34 and SA31EA33EA34 (all resolved to 1.74 Å; Harpaz *et al.*, manuscript in preparation). Both the structures of IV76 and SG31 share with truncated wild-type the problem of crystal contacts at the helix N-terminus. This is not the case, however, for the three multiple mutant structures, which pack in a different space group from wild-type, SG31 and IV76 mutants.

(iii) Reliability of the Available NMR and Crystal Structures of CI2:

A prerequisite to the rational design of mutants and correct interpretation of resulting data is the availability of a well-defined protein structure. It is important to know, therefore, how reliable available structural information is, particularly at the site of mutation. In this study of CI2, two regions of the protein are of special interest: the N-terminus of the α-helix (especially residues 31, 33 and 34) and the hydrophobic core (in particular residues 48, 68 and 76). Figures 1.3a - b and 1.4a - b compare the crystallographic B-factors (McPhalen & James, 1987) and NMR rms deviation from the average (Ludvigsen *et al.*, 1992) for both backbone and side chain atoms of residues 19 - 83 in wild-type CI2.

According to NMR (Ludvigsen *et al.*, 1992), the rms deviation from the average for the side chain atoms of helix residues Glu33 and Glu34 is high (around 3.0 Å), indicating that these side chains are likely to be flexible in solution. This compares with an rms deviation of only 1.1 Å for side chain atoms of Ser31, and between 0.5 - 1.5 Å for residues in the hydrophobic core. The crystal structures available for CI2, both free and in complex with subtilisin Novo (McPhalen *et al.*, 1985; McPhalen & James, 1987), have been refined with high B-factors for most solvent exposed atoms in the side chains of residues in the helix (where the B-factor is a measure of how localised an atom is in the unit cell and includes not only temperature-dependent vibrations of the atom but also static disorder within and between different unit cells). This is with the exception of the first residue in the helix, the N-cap Ser31. An average atomic B-factor of around 15 $Å^2$ obtained for both backbone and side chain atoms of Ser31 is indicative of an intermediately-defined residue. For residues Glu33 and Glu34, however, the B-factors for backbone and side chain atoms are high (around 20 $Å^2$ and 30-40 $Å^2$, respectively). Because residues 33 and 34 are both predominantly solvent exposed, they are probably free to occupy a wide-range of geometries. By comparison, the B-factor associated with the backbone and side chain atoms of hydrophobic core residues Leu68 and Ile76 are low (around 10 $Å^2$), indicating that these residues are well-defined in the crystal structure of CI2.

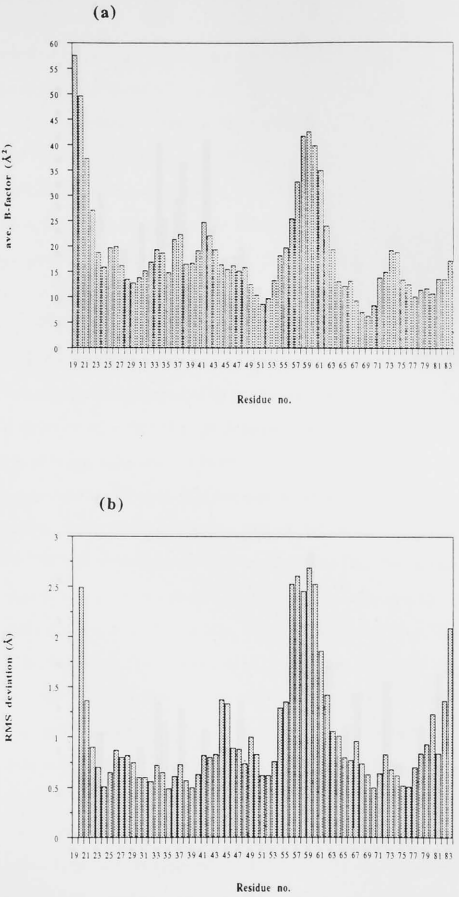


Fig. 1.3: Comparison of a) crystallographic average atomic B-factors with b) NMR rms deviation from the average for backbone atoms in the wild-type structure of CI2.

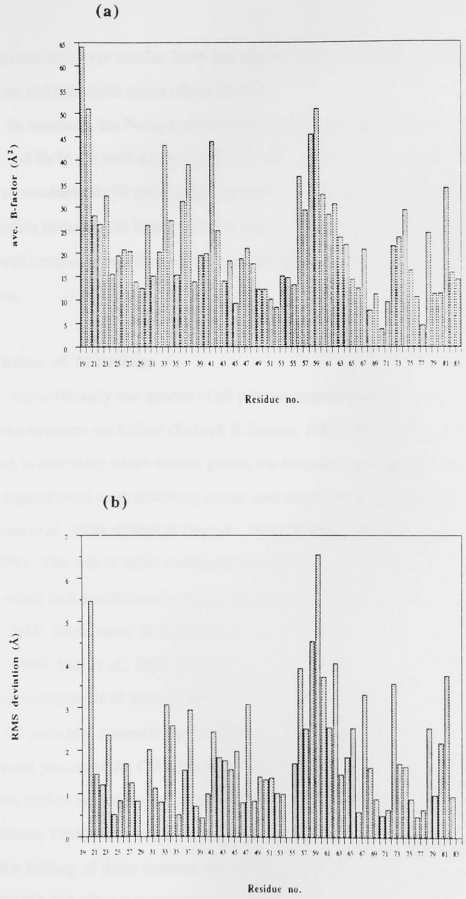


Fig. 1.4: Comparison of a) crystallographic average atomic B-factors with b) NMR rms deviation from the average for side chain atoms in the wild-type structure of CI2.

The hydrophobic core residue Ile48 has slightly higher B-factors associated with its backbone and side chain atoms (about 16 $Å^2$).

In summary, the N-cap residue Ser31 and the hydrophobic core residues Ile48, Leu68 and Ile76 are well-defined in both NMR and crystal structures, while the side chain of residues Glu33 and Glu34 are likely to be flexible in solution. Structural information pertaining to Glu33 and Glu34 (and similarly other surface residues) must be viewed cautiously in order to avoid over-interpretation of mutational effects on these residues.

1.6 Choice of Mutations

Approximately one quarter of all primary sequences found in globular proteins of known structure are helical (Kabsch & Sander, 1983). Much effort has been made in trying to determine which factors govern the formation and stability of helices, most of the experimental work involving amino acid substitutions in the middle of the helix (Altmann *et al.*, 1990; O'Neil & Degrado, 1990; Padmanaban *et al.*, 1990; Strehlow *et al.*, 1991). The role in helix stabilisation of the first and last four residues of the α-helix, which lack intrahelical hydrogen bond partners, has also been studied (Presta & Rose, 1988; Richardson & Richardson, 1988; Nicholson *et al.*, 1988; Serrano & Fersht, 1989; Bell *et al.*, 1991; Forood *et al.*, 1993). One of the first targets for site-directed mutagenesis of the truncated form of CI2 has been the N-terminus of the α-helix. Of particular interest is the first residue in the helix, the N-cap. Mutations in the N-terminal portion of the CI2 helix have been constructed with consideration given to previous studies on helices and helix stability (Serrano & Fersht, 1989; Richardson & Richardson, 1988). It is hoped that analysis of the thermodynamic and kinetic data from the folding of these mutants will further define the role of hydrogen bonding, electrostatic and other interactions in stabilising the N-termini of helices in CI2 and other proteins.

In addition, residues in the hydrophobic core of CI2 have been mutated in order to measure their contribution to the protein's stability. Formation of the hydrophobic

core is generally thought to be one of the major driving forces in protein folding (Dill, 1990a). Site-directed mutagenesis in combination with folding studies facilitates fine-structure analysis of the formation of these cores. In an attempt to better understand the effects of hydrophobicity on protein folding, correlations have been sought between the changes in stability and the environment of the mutated hydrophobic residue.

1.7 Aims of This Study

In the first instance, this thesis is concerned with equilibrium unfolding studies on mutants of CI2. In addition, the kinetics of unfolding and refolding of the same CI2 mutants are investigated to probe the nature of their unfolding and refolding pathways. The work presented here attempts to show that these CI2 mutants, like wild-type, adhere to a two-state model of protein folding, as characterised both thermodynamically and kinetically. Further to this, the effect of mutation at the α-helix N-terminus and hydrophobic core on the stability of CI2 is investigated. Finally, the transition state of the protein is partially characterised using the results obtained from thermodynamic and kinetic studies of the CI2 helix and hydrophobic core mutants.

Chapter 2

Materials and Methods

2.1 Chemicals

The buffer used in denaturation experiments was 2-(N-morpholino)ethanesulphonic acid (MES) and was purchased from Sigma. Sequenal grade GdnHCl was purchased from Pierce Chemicals. Water used in both equilibrium and kinetic experiments was purified to 15 Ω resistance by an Elgastat system. Ammonium sulphate was enzyme grade and purchased from BDH. All additional chemicals were purchased from Sigma, unless stated otherwise.

2.2 Media

Cells were grown in 2 x TY media (16 g L^{-1} tryptone, 10 g L^{-1} yeast extract, and 5 g L^{-1} NaCl in deionised water). Sterile 25 mL LB-Amp plates were made with 10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 10 g L^{-1} NaCl, 15 g L^{-1} bacto-agar, and 25 µL 5N NaOH in de-ionised water. Ampicillin was filter sterilised using a 0.22 µm sterile filter (Ministart NML, Sartorius) and was added to the autoclaved media (final concentration of 50 µg mL^{-1}) when the temperature of the media was less than 55 °C. Tryptone and yeast extract were purchased from Lab M, Bury U.K. Bacto-agar was purchased from Difco.

2.3 Bacterial Strains

Escherichia coli TG2 cells carry the gene for the F'-male plasmid encoding the F-pilus, which allows for infection with bacteriophage M13KO7 (see below). TG2

cells are recombination deficient. The genotype of TG2 cells is as follows: supE hsdΔ5 thi Δ(lac-proAB) Δ(srl-recA)306::Tn10(tet^r) F'[traD36 proAB$^+$ lacI^q lacZΔM15].

2.4 Recombinant CI2

A single copy of the truncated CI2 gene (expressing residues Met20 - Gly83) was subcloned into pTZ18U to create the high level expression vector, pCI2 (constructed by Michael Rheinnecker). pCI2 contains an ampicillin resistance gene, for selection purposes, and contains the bacteriophage f1 origin of replication, allowing for production of single stranded DNA upon infection with helper phage M13KO7 (see section 2.7). The structural gene is under the control of the lac promoter. The original truncated gene was obtained from F. Poulsen, Carlsberg Laboratory, Denmark.

2.5 Oligonucleotide Design

Mismatched oligonucleotide primers for Sayers & Eckstein site-directed mutagenesis (see below) were designed to have a melting temperature of approximately 30 °C on each arm around the mismatch. It was assumed that the G or C nucleotides contributed 4 °C and A or T nucleotides contributed 2 °C towards stabilising the double stranded form. This ensured that the oligonucleotide was not displaced during the primer extension step in the mutagenesis reaction. The synthetic mutating oligonucleotide was designed to be complementary to the single-stranded plasmid form. An oligonucleotide constructed in this way could contain up to three mismatches.

Mismatched oligonucleotides were also used as coding-strand primers in 'Inverse' PCR (IPCR) mutagenesis. A non-coding strand primer was designed to anneal 'back-to-back' with the coding strand primer. All oligonucleotides were synthesised on an Applied Biosystems 380B DNA Synthesiser. See Table 2.1 for a list of mutant oligonucleotides that were synthesised.

Table 2.1: Oligonucleotides Used in Site-Directed Mutagenesis of CI2

Mutant	Oligonucleotide Sequence (5'→ 3')	Mismatched bases
SG31	TCCTCCAC**CCC**TTTCCCCAC	**CGA**[a]
SA31	TCCTCAC**CGC**TTTCCCCAC	**CGA**
EQ33	TGGCCTC**CTG**CACCGATTTC	**CTC**
ED33	CTTGGCCTC**ATC**CACCGATT	**CTC**
EN33	CTTGGCCTC**ATT**CACCGATT	**CTC**
EQ34	TTCTTGGC**CTG**CTCCACCGA	**CTC**
ED34	CCTTCTTGGC**GTC**CTCCACC	**CTC**
EN34	CCTTCTTGGC**ATT**CTCCACC	**CTC**
IV48	GCAGAACTAT**GAC**TTGCGCCTC	**GAT**
IA48	CAGAACTAT**GGC**TTGCGCCTC	**GAT**
LA68	TATCGACAAA**GGC**GCGGACGC	**GAG**

IPCR Mutants	IPCR Oligonucleotide Sequence	Mismatched bases
EA33EA34	CTTCTTGGC**CGCCGC**CACCGATTTCCCCAC	**CTCCTC**
SG31EA33EA34	CTTCTTGGC**CGCCGC**CAC**CCC**TTTCCCCAC	**CTCCTC**xxx**CGA**
SA31EA33EA34	CTTCTTGGC**CGCCGC**CAC**CGC**TTTCCCCAC	**CTCCTC**xxx**CGA**
Inverse Primer	GTGATTCTGCAGGACAAGCCAGAGGCGCAA	No mismatches

[a] nucleotides in bold type-face code for the mutated amino acid
N.B. The mutant IA48IV76 was not intentionally made, but was a random mutant found during sequencing of the other hydrophobic core mutants, and hence no oligonucleotide was made to code for this mutation.

2.6 Mutagenesis

Site-directed mutagenesis (SDM), employing the method developed by F. Eckstein (Sayers & Eckstein, 1988) was carried out using the kit produced by Amersham, U.K., according to the manufacturer's instructions. Construction of the hydrophobic core mutants IV48, IA48, and LA68 was kindly performed by Dr Marco Moracci.

Briefly, a synthetic oligonucleotide (oligo) containing the mutated sequence was hybridised against the single-stranded plasmid. The Klenow fragment of DNA *polymerase I* was used to extend the oligo, incorporating a thio-analogue of the nucleotide-triphosphate, dCTPαS. This allowed for the selection of the mutant strand of the heteroduplex form of the plasmid. The heteroduplex (one mutant and one wild-type strand) was treated with the restriction enzyme, *NciI*, which nicked the strand that did not contain the thio-analogue, i.e., the wild-type strand. The nicked wild-type strand was degraded, 3' → 5', by the *E. coli Exonuclease III* (*Exo III*). After inactivation of the *Exo III* upon heating, a heteroduplex of the mutant plasmid was produced by the addition of *DNA polymerase I*.

Mutant plasmid was then transformed into *E. coli* TG2 cells (see recombinant CI2 expression and purification for method). Mutants were screened directly by performing dideoxy sequencing (see below). The efficiency of mutagenesis was between 75 - 90%.

Mutants were also made using IPCR mutagenesis, as described by Clackson and coworkers (Clackson *et al.*, 1991). The mutants EA33EA34, SG31EA33EA34, and SA31EA33EA34 were constructed by Michael Rheinnecker using IPCR mutagenesis.

Briefly, in this procedure, mutants in the plasmid were directed by one of two apposed primers which annealed "back-to-back". PCR resulted in the amplification of the entire plasmid with the site-directed mutation incorporated. This product was then purified and self-ligated to regenerate a circular vector. The procedure was extremely rapid and 50-100% of the resulting colonies harboured the mutant sequence, as determined by dideoxy sequencing.

2.7 Screening of Mutant Plasmids by Dideoxy Sequencing

Dideoxy sequencing of site-directed mutant plasmids was carried out according to the method developed by Sanger (Sanger *et al.*, 1977) using a kit purchased from Sequenase, U.K., following the manufacturer's instructions.

The single-stranded DNA template used for this method was produced by infection of TG2 cells harbouring the pCI2 plasmid with M13KO7 helper phage (Vieira and Messing, 1987), according to the method outlined by Maniatis. M13KO7, a male-specific bacteriophage, infects its host via the F-pilus, and produces single-stranded DNA via the rolling circle method of replication. It is important that the M13KO7 helper phage, used to propagate single-stranded pCI2, be grown from fresh plaques. M13KO7 phage, which when stored at 4 °C will remain viable for more than one year, was purchased from Pharmacia Molecular Biology Division.

2.8 Recombinant CI2 Expression and Purification

The following expression and purification methods were used for both wild-type and mutant CI2. pCI2 was transformed into *E. coli* TG2 cells, using the $CaCl_2$ method as described by Maniatis *et al.* (Molecular Cloning: A Laboratory Manual, C.S.H. Laboratory Press). These cells were then plated onto L-amp plates and a single colony was picked (TG2 cells harbouring the plasmid have ampicillin resistance conferred on them). This colony was then used to inoculate a starter culture grown in 2 x TY medium containing ampicillin (50 $\mu g\ mL^{-1}$). The culture was grown overnight and was used to inoculate (1:100 dilution) a larger volume of 2 x TY containing ampicillin (50 $\mu g\ mL^{-1}$). CI2 expression was induced by the addition of IPTG (50 mM) when the absorbance was between 0.4 - 0.8 at 600 nm, although it had been found previously that the level of CI2 expression was independent of induction time (Campbell, 1988). CI2 is expressed in the cytoplasm. The culture was grown overnight at 37°C in a Model G25 Incubator Shaker, New Brunswick Scientific Co. Inc., New Jersey, with vigorous shaking (320 rpm) to an absorbance of 6 - 8 at 600 nm (approximately overnight).

The cells were harvested and concentrated using an RC5C Sorvall Instruments Centrifuge, with a fixed angle GS-3 rotor, set at 5,000 rpm for 15 minutes at 4 °C. The resulting cell paste was resuspended in a small volume of pre-cooled 10 mM Tris-HCl pH 8.0, 1 mM EDTA. The resuspended cells were sonicated for 6 x 30 seconds on maximum power at 4 °C using a Heats Systems W-373 Ultrasonic Sonicator, 12.5 mm tip. The sonicator was previously tuned against air. The sonication step ensured that the cell membrane was broken and that intracellular proteins were released into the supernatant. Sufficient time was allowed in between each sonication to ensure that the cell suspension remained at 4°C. An RC5C Sorvall Centrifuge with a fixed angle SS-34 rotor was set at 15,000 rpm for 20 minutes at 4 °C to remove cell debris. In order to precipitate unwanted proteins, ammonium sulphate was slowly added to the supernatant to a final concentration of 40 % (w/v) (23 g per 100 mL supernatant), ensuring that all ammonium sulphate lumps were broken-up before addition. This mixture was stirred gently at 4 °C for 20-30 minutes in order that all ammonium sulphate had been dissolved. The precipitated protein was pelleted using an SS-34 rotor set at 18,000 rpm for 40 minutes at 4 °C. The supernatant was transferred to new containers and ammonium sulphate was again added, to a final concentration of 80 % (w/v) (26 g per 100 mL supernatant) so as to precipitate CI2. After slow addition of the ammonium sulphate, the mixture was gently stirred for at least one hour (or overnight) at 4 °C. The precipitated protein was pelleted using an SS-34 rotor, as previously described. The protein pellet was resuspended in as small a volume as possible of 50 mM Tris-HCl pH 8.6, and was then dialysed extensively against 50 mM Tris-HCl pH 8.6. Benzoylated cellulose tubing with a M_r cut-off of 2,000 D was used for all dialysis. The tubing, which was stored at 4 °C in a solution of 0.002% 1-hydroxypyridine-2-thione as a preservative, was washed thoroughly in de-ionised water before use. Irreversibly precipitated proteins were removed by centrifugation using an SS-34 rotor, as described before. DEAE-Trisacryl was pre-equilibrated in 50 mM Tris-HCl pH 8.6, and was added to the protein solution and the mixture stirred (about 2 mL of DEAE-Trisacryl was added to 50 mL of protein solution). The supernatant containing CI2

was decanted, and any residual DEAE-Trisacryl was removed by centrifugation using an IEC Centra-4X Centrifuge, Bedfordshire, U.K., set at 2,500 rpm for 5 minutes. Nucleic acid contaminants were precipitated with protamine sulphate; 1 mg of protamine sulphate (in water) was added for every 30 mg of protein. The protamine sulphate was first dissolved in de-ionised water and the pH adjusted to 7.0 by the addition of a few microlitres of 10 mM NaOH. Precipitated nucleic acid was pelleted using an SS-34 rotor set at 18,000 rpm for 10 minutes at 4 °C.

The protein sample was loaded onto a HiLoad Superdex G75 prep. grade FPLC gel filtration column (26 mm x 60 cm), purchased from Pharmacia, Sweden. The column had been pre-equilibrated at a flow rate of <2.6 mL min^{-1} with 50 mM Tris-HCl, pH 8.6, and 150 mM NaCl to prevent non-specific binding of proteins to the column. A maximum of 13 mL of protein solution was loaded onto the column, which was run at a flow rate of 1.3 mL min^{-1} with the chart recorder speed at 0.25 cm min^{-1}. 4 mL fractions were collected, with CI2 eluting as a single, well resolved peak with an elution volume of approximately 210 mL. Fractions were collected and assayed for inhibitory activity, as described below, and those fractions containing CI2 were pooled. Purified CI2 was exhaustively dialysed against de-ionised water to remove salt. The protein was then flash frozen and stored at -70 °C. The purified protein was judged homogeneous by NaDodSO_4-polyacrylamide gel electrophoresis (Laemmli, 1970). Typical yields for both wild-type and mutant proteins were 10 - 60 mg L^{-1}.

2.9 Qualitative Microtitre Plate Assay for CI2

The qualitative assay for the presence of CI2 was based upon the ability of CI2 to inhibit the subtilisin-catalysed hydrolysis of the synthetic substrate, succinyl-Ala-Ala-Pro-Phe-*para*-nitroanilide (sAAPFpNA). In the absence of any inhibition, subtilisin catalyses the hydrolysis of sAAPFpNA and releases *para*-nitroaniline, causing the solution to turn yellow in colour. When CI2 is present, however, such hydrolysis is prevented, and so the solution remains clear. The assay was carried out as follows.

5 - 10 μL of fraction (thought to contain CI2) was pipetted into a microtitre plate well (Sero-Wel, Sterilin). To this was added approximately 2 μL of 1 mM subtilisin solution, mixing well using a P-20 Gilson Pipetman. sAAPFpNA was diluted from a 50 mM stock solution (in dry DMSO) into 0.1 M Tris-HCL pH 8.6 to a final concentration of 0.5 mM. 50 μL of this solution was added to the microtitre plate well and was left to incubate at room temperature for 5 - 10 minutes. The solution in wells that contained active subtilisin turned yellow, upon release of *para*-nitroaniline. Those wells that contained fractions with active CI2 remained clear due to the effective inhibition of subtilisin by CI2. If the results of the assay were unclear, then it was repeated using varying amounts of enzyme or fraction, until ambiguity was removed. The assay was used to test fractions collected from the Superdex G75 gel filtration column during the final stages of CI2 purification.

2.10 Nuclear Magnetic Resonance

The two-dimensional structure of the truncated form of CI2 was verified (by Dr Sophie Jackson, University Chemical Laboratories, Cambridge) using a Bruker AMX 500 Nuclear Magnetic Resonance Spectrometer. A 500 μL sample of 2 - 3 mM CI2, pH 4.5, was used. 1-D ^{1}H-NMR experiments were also performed by Dr Jackson (on long-form CI2) to verify that the 2-state folding transition monitored fluorimetrically was due to global denaturation of the protein and not caused by some local folding event in the proximity of the tryptophan. The latter experiments were carried out at various GdnHCl concentrations, at 25 °C.

2.11 N-terminal Protein Sequencing

N-terminal sequencing of the truncated form of CI2 (0.25 mg mL^{-1}) was performed (by Dr Len Packman, Biochemistry Dept., Cambridge University) using an Applied Biosystems 477A Protein Sequencer.

2.12 Calculation of Solvent Accessible Surface Area of Mutants

The difference in solvent accessible surface that is buried on folding between wild-type and mutant is calculated in the following way. The solvent accessible surface area of the residue in the wild-type protein is calculated using the Lee & Richards (1971) algorithm. This value is then subtracted from the solvent accessible surface area for that residue in a tripeptide (Miller *et al.*, 1987). This gives a value *A*. In parallel, a theoretical mutation was performed by deleting the appropriate methylene group(s) from the co-ordinate file, and the solvent accessible area of the mutated residue in the protein was calculated. This was subtracted from the value calculated for the mutated residue in a tripeptide to give a value *B*. The difference between the two, *A-B*, is the loss of solvent accessible surface area buried in the folded protein upon mutation. All solvent accessible surface area calculations were performed by Dr Sophie Jackson, with the help of Dr Cyrus Chothia.

2.13 Fluorescence Spectroscopy

Fluorescence spectroscopy was used in both GdnHCl-induced denaturation and kinetic experiments. The intrinsic fluorescence of CI2 increases upon denaturation, due to the exposure of a tryptophan that is normally buried in the folded protein. This allows the unfolding and refolding of CI2 to be monitored using fluorescence spectroscopy. It has been found that the maximal change in fluorescence upon denaturation occurs at an excitation wavelength of 280 nm and an emission wavelength of 356 nm.

The fluorescence yield at 356 nm of the unfolded and folded forms of the protein at equilibrium is constant with respect to GdnHCl concentration, up to 5.6 M only, contrary to the findings of Schmid (1989). Control experiments performed with Trp alone and Trp-containing barnase demonstrate that the intrinsic fluorescence of Trp is linear until 5.6 M GdnHCl, after which the fluorescence increases with increasing denaturant concentration.

2.14 Guanidinium Chloride Denaturation Experiments

GdnHCl solutions were prepared gravimetrically using volumetric flasks. The solutions were divided into 800 µL aliquots using an SMI positive displacement pipetter (Americal Hospital Supply) with a repetitive pipetting attachment and were stored at -20 °C. 100 µL of CI2 stock solution in 450 mM MES pH 6.3 was diluted into 800 µL of the appropriate denaturant concentration, using an SMI positive displacement pipetter. The final concentration of CI2 was approximately 2.5 µM. This was repeated for every data point in the denaturation experiment. Typically 30 - 40 points were measured during an experiment, from 0 - 5.6 M GdnHCl, for wild-type and mutant CI2 proteins. Solutions containing protein and denaturant were pre-equilibrated at 25 °C for approximately 1 hour. Spectroscopic measurements were conducted using a thermostatted cuvette holder at 25 °C. The temperature was monitored throughout the experiment using a thermocouple immersed in a neighbouring cuvette in the cell holder. Unfolding has been found to be completely reversible (Jackson & Fersht, 1991a) with over 95% of the CI2 refolded to an active, native conformation. SMI positive pipetters and Hamilton syringes were used for all manipulations. A Perkin-Elmer LS5B Luminescence spectrometer was used, with slit-widths set at 2 nm.

2.15 Differential Scanning Microcalorimetry

Measurements were carried out in 10 mM Glycine-HCl buffer at pH 3.0 for all mutants, except the mutants SG31, IV48, IA48, LA68 and IA48IV76, for which measurements were performed in 10 mM Formate buffer at pH 3.5. Under these sets of conditions, the protein was soluble and no aggregation was observed on heating. Protein solutions were exhaustively dialysed against buffer at 4 °C, and centrifuged at 13,000 rpm in an MSE microcentrifuge for 5 mins. before measurements were taken. The final dialysis buffer was used for reference and baseline scans. All solutions were degassed under vacuum with gentle stirring for 2 - 3 mins. before loading to avoid bubble formation at higher temperatures.

Protein concentrations in solution were determined by absorbance measurements with a correction for light scattering (Winder & Gent, 1971), using a Cary 3 U.V.-Visible Photospectrophotometer, Varian, U.S.A. Absorbances were measured over the range of 320 - 365 nm, at 5 nm intervals; over this wavelength range the contribution to absorbance is due to light scattering only and not the presence of protein. The logarithm of the absorbances measured over this range were plotted against the wavelength, with data fitting to a straight line. The data was then extrapolated to 282 nm to calculate the contribution to the measured absorbance at this wavelength due to light scattering. For different protein solutions, this was found to be around 0.005 - 0.015 absorbance units at 282 nm . Prior to accurate calculation of protein concentration, this value was subtracted from the measured absorbance at 282 nm. The concentration of protein used for each calorimetry experiment was 0.5 - 1.5 mg mL^{-1}. The extinction coefficient calculated for the truncated form of CI2, 6965 M^{-1} cm^{-1} at 282 nm, was used for all protein concentration calculations. Extinction coefficients were also measured for several mutants but the value was found to vary by less than 2% (S. Jackson, personal communication) and so the wild-type extinction coefficient was used throughout.

A Microcal MC-2 Differential Scanning Microcalorimeter (Microcal Inc., U.S.A.) was used to perform calorimetric measurements. The machine was equipped with tantalum cells of 1.2163 mL volume. Measurements were performed under a constant pressure of 2 bar at a notional scan-rate of 1 K min-1, for 60 minutes, with constant power being applied during the scan. Several baseline scans were performed prior to calorimetric measurement of protein samples being made. In order to check the reversibility of the thermal unfolding, samples were rescanned after cooling to below the transition temperature over a period of 50 mins. For a detailed account of DSC methodology, reference should be made to Cooper & Johnson (1992).

Microcal Origin software, U.S.A., was used for analysis and fitting of DSC thermograms as single peak non-two-state transitions. Thermograms were corrected by subtraction of buffer baselines. Using a molecular weight of 7308 Da and cell volume

of 1.2163 mL for all samples, corrected thermograms were then converted to excess heat capacity. In order to eliminate changes in heat capacity between folded and unfolded states of the protein, an interpolative baseline between pre- and post-transition baseline slopes in proportion to the progress through the transition was used. Calorimetric measurements on CI2 wild-type and hydrophobic core mutants were performed by Dr Chris Johnson at the IRC for Protein Engineering, Cambridge, U.K.

2.16 GdnHCl-Jump Kinetic Experiments

Both unfolding and refolding kinetic reactions of CI2 were followed using a Perkin-Elmer MPF-44B fluorescence spectrophotometer equipped with a rapid mixing head. Unfolding of the protein was initiated by dilution of the aqueous protein solution (approximately 45 μM in 50 mM MES pH 6.3), 1:10, into denaturing concentrations of GdnHCl. The final concentrations of GdnHCl were between 3.0 and 7.0 M, in 50 mM MES pH 6.3. After pre-equilibration of the protein in 6.5 M GdnHCl, 50 mM MES pH 6.3, refolding was initiated by diluting the protein solution, 1:10, into low concentrations of GdnHCl. Final GdnHCl concentrations were between 0.59 - 4.0 M, in 50 mM MES pH 6.3.

Using two Hamilton syringes, the solutions were driven through the mixing chamber manually resulting in a mixing ratio of 1:10. A T-jet mixing chamber was contained in the mixing device, followed by a 30 ms delay loop to ensure complete mixing of the solutions before observation. A Hellma flow-through compact cell (10 mm x 3 mm x 3 mm) was used. Both the observation cell and reservoir syringes were thermostated to 25 °C separately, using two cooling-water baths. The temperature of each was monitored using an Edal Instrument Thermometer, with temperatures kept to within ± 0.1 °C. The excitation and emission wavelengths were set at 280 nm and 356 nm, respectively, with slit widths of 12 - 15 nm for each experiment. Data were acquired using a Tandon Microcomputer, a DT2801 Data Translation board and the Bio-kine data acquisition software package by Bio-logic.

Unfolding data were fitted to a single exponential rate equation, allowing for offset and drift (equation 2.1):

$$F(t) = A_0\{1 - \exp(-k_u t)\} - mt + C \qquad (2.1)$$

where $F(t)$ is the observed fluorescence at time t, A_0 the amplitude, k_u the rate constant of unfolding, m the slope of the drift (from baseline instability), and C the offset. A large increase in fluorescence is observed at 356 nm when CI2 unfolds. Fig. 2.1 represents a typical unfolding trace for CI2.

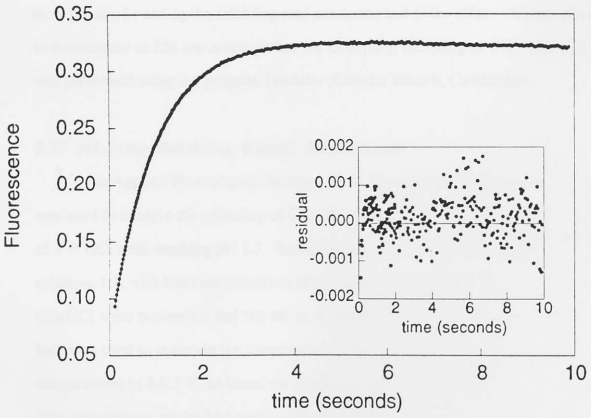


Fig. 2.1: A Typical Unfolding Trace for CI2 (inset : residual error)

Refolding is a triphasic process with a very fast, major phase (77% of the signal) and two slow, minor phases (Jackson & Fersht, 1991b). Over the time-scale during which the refolding reaction was monitored (typically from 0 - 500 ms), the slowest phase was found to be so slow with respect to the first two phases that it did not affect the reaction appreciably (Jackson & Fersht, 1991b). The rate constant for the fast phase is greater than 2000 times faster than that for the slowest phase. Thus, the fast and intermediate phases were fitted to a double exponential decay rate equation (allowing for offset but not drift), ignoring the slowest phase, as in equation 2.2:

$$F(t) = A_1\exp(-k_1t) + A_2\exp(-k_2t) + C \qquad (2.2)$$

where $F(t)$ is the fluorescence at time t, A_1 and A_2 are the amplitudes of the fast and slow phases, k_1 and k_2 the refolding rate constants, and C the offset. A large decrease in fluoresence at 356 nm accompanies the Gdn-jump refolding of CI2. Data analysis was performed using the program Enzfitter (Elsevier Biosoft, Cambridge).

2.17 pH-Jump Refolding Kinetic Experiments

An Applied Photophysics Stopped-Flow Spectrophotometer Model SF 17MV was used to observe the refolding of CI2. Protein (≈20μM) in water was denatured by addition of 5 N HCl until reaching pH 1.7. Refolding was then initiated by mixing (2 ms dead-time) the protein solution, 1:1, with low concentrations of guanidinium chloride. Final concentration of GdnHCl were between 0 and 0.6 M, in 50 mM MES pH 6.3. A Grant LTD6 water bath was used to maintain the temperature of the cell and reservoir syringes, keeping temperatures to ± 0.1 °C as based on an internal probe in the stopped-flow apparatus. This temperature probe had previously been calibrated against an Edale Instrument Thermometer. A Radiometer PHM64 Research pH Meter was used for pH measurements.

Refolding data from pH-jump experiments were treated in an identical manner as refolding data obtained from GdnHCl-jump experiments, as in equation 2.2. A large

decrease in fluorescence at 356 nm accompanies the pH-jump refolding of CI2, where Fig. 2.2 represents a typical pH-jump refolding trace for CI2. The SF 17MV software package by P.J. King, Applied Photophysics was used for fitting pH-jump refolding data.

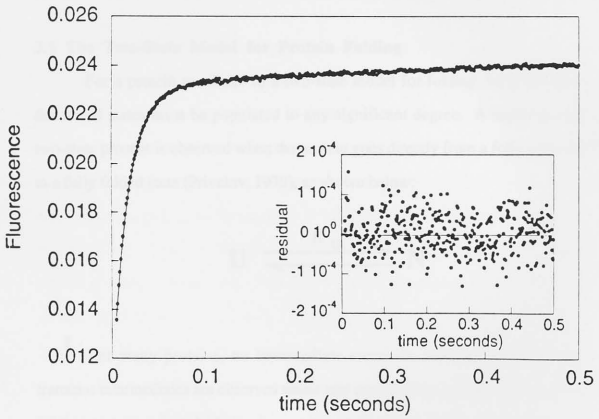


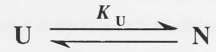
Fig. 2.2: A Typical pH-jump Refolding Trace for CI2
(inverted scale; inset is residual error)

Chapter 3

Mutants of CI2 Fit the Two-State Model for Protein Folding

3.1 The Two-State Model for Protein Folding

For a protein to adhere to a two-state model for folding, only the native and denatured states must be populated to any significant degree. A highly cooperative two-state process is observed when the protein goes directly from a fully unfolded state to a fully folded state (Privalov, 1979), as shown below:



For many proteins, no intermediate states are detected at equilibrium, but transient intermediates are observed under non-equilibrium conditions along the kinetic folding pathway, thus deviating from a two-state model. This is the case for barnase, where a two-state process is observed at equilibrium only; an intermediate state is detected along the refolding pathway of that protein (Matouschek *et al.*, 1990; Bycroft *et al.*, 1990). By comparison, CI2 is a rare example of a protein that adheres to a two-state model for folding under both equilibrium and non-equilibrium conditions (Jackson & Fersht, 1991a, 1991b).

In order to ascertain whether this is also the case for mutants of CI2, the thermodynamics and kinetics of folding of these mutants are studied. The thermodynamics of CI2 unfolding are studied using both GdnHCl-induced equilibrium denaturation (monitored by fluorescence) and thermal denaturation (measured by

differential scanning microcalorimetry). The effect of mutation on CI2 stability is measured, as well as the adherence to a two-state transition at equilibrium. Kinetic studies are performed on both wild-type and mutant protein. Using GdnHCl-jump and pH-jump stopped-flow experiments, changes in the fluorescence of CI2 upon unfolding and refolding are observed and rate constants are measured as a function of [GdnHCl]. The kinetic data is then fitted to a two-state model in order to determine whether CI2 mutants adhere to this model kinetically as well as thermodynamically.

3.2 Analysis of GdnHCl-Induced Denaturation Data

(i) Calculation of $[GdnHCl]_{50\%}$, $m_{U\text{-}F}$ and $\Delta G_{U\text{-}F}$:

The data obtained from the GdnHCl-induced denaturation of wild-type and mutant CI2 are analysed according to a two-state model of protein folding, where only the native and denatured states are populated to any significant degree. A typical example of the change in fluorescence of CI2 wild-type upon titration with GdnHCl is shown in Fig. 3.1a, where a sharp two-state transition between fully folded and unfolded protein can be seen. Figs. 3.1b - p show the change in fluorescence of CI2 mutants upon titration with GdnHCl. These represent typical curves since experiments were repeated more than once for each mutant.

Based on fluorescence measurements taken over a range of GdnHCl concentrations it is possible to calculate both the equilibrium constant for unfolding, K_U, and the free energy of unfolding, $\Delta G_{U\text{-}F}$, from equation 3.1:

$$K_U = \frac{(F_N - F)}{(F - F_U)} = \exp(-\Delta G_{U\text{-}F}/RT) \qquad (3.1)$$

where F is the observed fluorescence, F_N is the fluorescence of the native state, F_U is the fluorescence of the denatured state of the protein, and R is the gas constant, 1.987 cal mol^{-1} K^{-1} (8.314 J mol^{-1} K^{-1}). It has been previously stated that the values of F_N and F_U are independent of [GdnHCl] up to 6.5 M (Schmid, 1989). However, it has now been shown that such independence only holds true for concentrations up to 5.6 M

GdnHCl (see Chapter 2, section 2.13). Provided fluorescence measurements are made below this value, equation 3.1 can be used directly in order to calculate K_U and $\Delta G_{U\text{-}F}$. Tanford and others (Tanford, 1968; Pace, 1986) have made the empirical observation that the free energy of protein unfolding in the presence of denaturant, such as GdnHCl, can be related linearly to the concentration of that denaturant, as in equation 3.2:

$$\Delta G_{U\text{-}F}^{[GdnHCl]} = \Delta G_{U\text{-}F}^{H_2O} - m_{U\text{-}F}[GdnHCl] \qquad (3.2)$$

where $\Delta G_{U\text{-}F}^{[GdnHCl]}$ is the free energy of unfolding at a particular denaturant concentration, $\Delta G_{U\text{-}F}^{H_2O}$ is the free energy of unfolding in water, and $m_{U\text{-}F}$ is the slope of the transition region and is related to the increase in degree of exposure of the protein to solvent upon denaturation. If it is assumed that both F_N and F_U are independent of [GdnHCl] between 0 - 5.6 M, then the following equation (3.3) is obtained, from a combination of equations 3.1 and 3.2:

$$F = F_N - (F_N - F_U)\frac{\exp((m_{U\text{-}F}[GdnHCl] - \Delta G_{U\text{-}F}^{H_2O})/RT)}{1 + \exp((m_{U\text{-}F}[GdnHCl] - \Delta G_{U\text{-}F}^{H_2O})/RT)} \qquad (3.3)$$

However, if it is assumed that both F_N and F_U are linearly dependent on [GdnHCl], i.e., $F_N = \alpha_N + \beta_N[GdnHCl]$ and $F_U = \alpha_U + \beta_U[GdnHCl]$, then equation 3.4 may be used instead:

$$F = \frac{(\alpha_N+\beta_N[GdnHCl]) + (\alpha_U+\beta_U[GdnHCl])\exp((m_{U\text{-}F}[GdnHCl] - \Delta G_{U\text{-}F}^{H_2O})/RT)}{1 + \exp((m_{U\text{-}F}[GdnHCl] - \Delta G_{U\text{-}F}^{H_2O})/RT)} \qquad (3.4)$$

where α_N and α_U are the intercepts and β_N and β_U are the slopes of the baselines at low (i.e., native) and high (i.e., denaturing) GdnHCl concentrations, respectively. Values for $\Delta G_{U\text{-}F}^{H_2O}$ and $m_{U\text{-}F}$ can be calculated, with their standard errors, for both wild-type and mutant proteins using the above analysis.

When comparing the stability of wild-type and mutant proteins it is often more useful to know the error associated with the concentration of denaturant at which the mid-point of unfolding occurs, i.e., $[GdnHCl]_{50\%}$ (where $K_U = 1$), and the accuracy of estimates of $\Delta G_{U\text{-}F}^{[GdnHCl]}$, the free energy of unfolding at various concentrations of denaturant. At $[GdnHCl]_{50\%}$, equations 3.5 and 3.6 follow on from equation 3.2:

$$\Delta G_{U\text{-}F}^{H_2O} = m_{U\text{-}F}[GdnHCl]_{50\%} \qquad (3.5)$$

$$\Delta G_{U\text{-}F}^{[GdnHCl]} = m_{U\text{-}F}([GdnHCl]_{50\%} - [GdnHCl]) \qquad (3.6)$$

From equations 3.4 and 3.6, it is possible to obtain equation 3.7:

$$F = \frac{(\alpha_N+\beta_N[GdnHCl]) + (\alpha_U+\beta_U[GdnHCl])\exp((m_{U\text{-}F}([GdnHCl]-[GdnHCl]_{50\%}))/RT)}{1 + \exp(m_{U\text{-}F}([GdnHCl]-[GdnHCl]_{50\%})/RT)} \qquad (3.7)$$

Fluorescence values obtained from the GdnHCl-induced equilibrium denaturation experiment were fitted with a non-linear regression procedure, using equation 3.7 and applying the general curve fit option in the program *Kaleidagraph* (version 2.1.1, Synergy Software, PCS Inc., U.S.A.). This procedure results in errors for $[GdnHCl]_{50\%}$ that are significantly smaller than those obtained for either $\Delta G_{U\text{-}F}^{H_2O}$ or $m_{U\text{-}F}$ since a shorter extrapolation is used. Small errors in $m_{U\text{-}F}$ can lead to much larger errors in estimates of $\Delta G_{U\text{-}F}^{H_2O}$ due to the long extrapolation from the transition region of unfolding (within the range 2 - 4 M GdnHCl) to water (i.e., 0 M GdnHCl). From repeated experiments on wild-type and mutant proteins it is observed that the measurement of $[GdnHCl]_{50\%}$ is very reproducible, with identical results being obtained from data fitted to equations 3.3, 3.4 and 3.7.

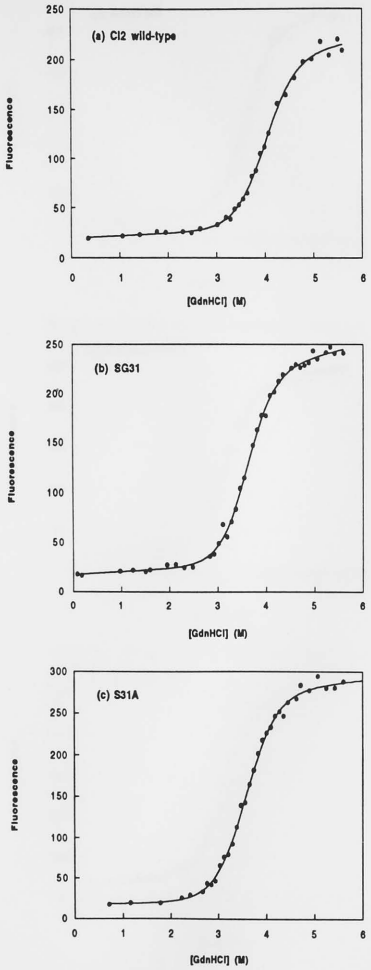


Fig. 3.1 a - c: GdnHCl-Induced Equilibrium Denaturation Curves for a) CI2 Wild-Type, b) SG31 and c) SA31

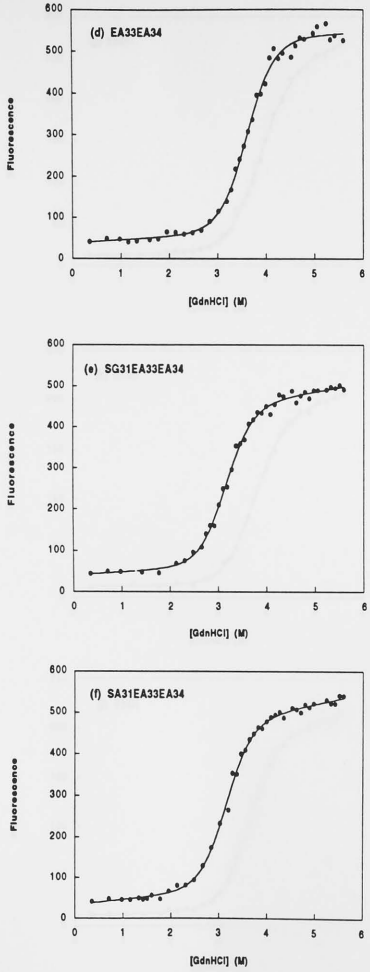


Fig. 3.1 d - f: GdnHCl-Induced Equilibrium Denaturation Curves for d) EA33EA34, e) SG31EA33EA34 and f) SA31EA33EA34

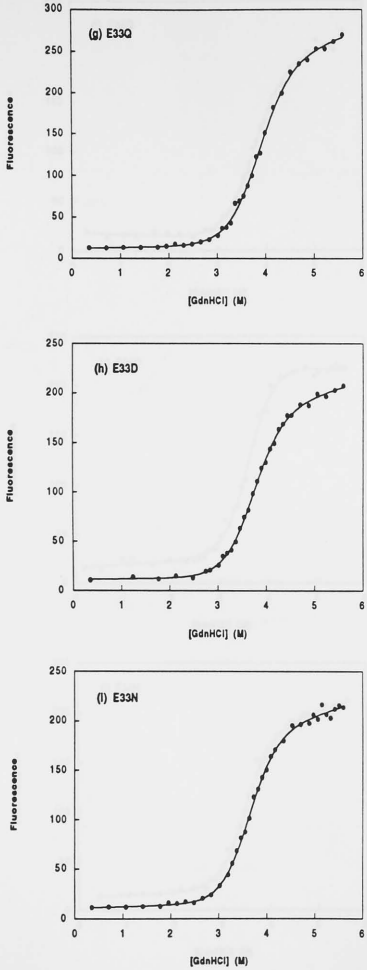


Fig. 3.1 g - i: GdnHCl-Induced Equilibrium Denaturation Curves for g) EQ33, h) ED33 and i) EN33

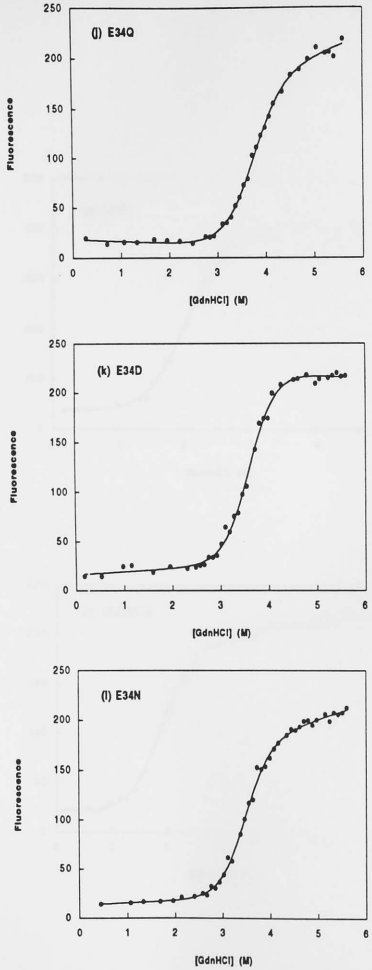


Fig. 3.1 j - l: GdnHCl-Induced Equilibrium Denaturation Curves for j) EQ34, k) ED34 and l) EN34

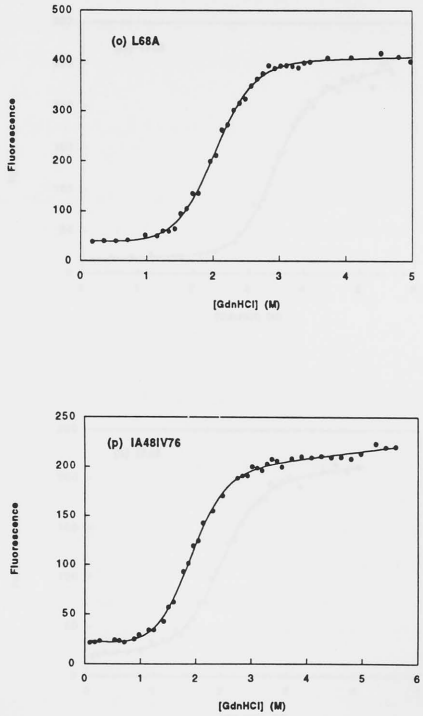


Fig. 3.1 m - n: GdnHCl-Induced Equilibrium Denaturation Curves for m) IV48 and n) IA48

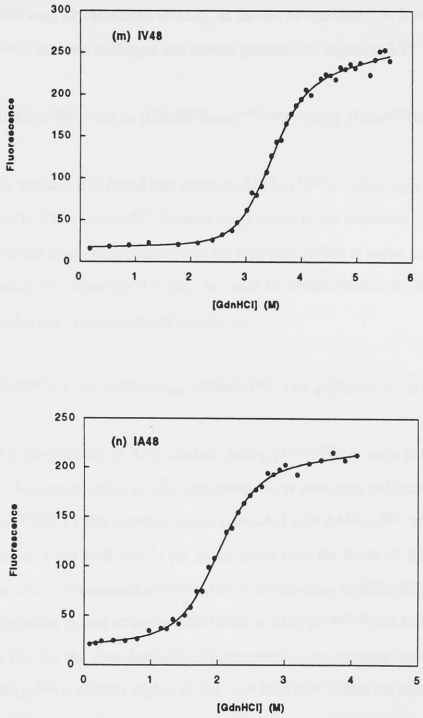


Fig. 3.1 o - p: GdnHCl-Induced Equilibrium Denaturation Curves for o) LA68 and p) IA48IV76

(ii) Calculation of $\Delta\Delta G_{U\text{-}F}$:

$\Delta\Delta G_{U\text{-}F}$ can be calculated in three different ways, although the error obtained for each value varies with the method used (see Table 3.1). In the first instance, $\Delta\Delta G_{U\text{-}F}{}^{H_2O}$ may be calculated directly, as shown in equation 3.8, from the difference in $\Delta G_{U\text{-}F}{}^{H_2O}$ between wild-type and mutant protein (see equation 3.5):

$$\Delta\Delta G_{U\text{-}F}{}^{H_2O} = (m_{U\text{-}F}\ [GdnHCl]_{50\%})^{\text{wild-type}} - (m_{U\text{-}F}\ [GdnHCl]_{50\%})^{\text{mutant}} \quad (3.8)$$

Using this method, it is found that errors in $\Delta\Delta G_{U\text{-}F}{}^{H_2O}$ are often very high (see Table 3.1; typically $\pm$ 0.5 kcal mol^{-1}) because large errors in the estimates of $\Delta G_{U\text{-}F}{}^{H_2O}$ can arise due to the long extrapolation from the transition region to water, as already stated.

Secondly, equation 3.9 may be used to obtain values of $\Delta\Delta G_{U\text{-}F}{}^{[GdnHCl]}$ determined at any concentration of denaturant:

$$\Delta\Delta G_{U\text{-}F}{}^{[GdnHCl]} = \{m_{U\text{-}F}([Gdn]_{50\%} - [Gdn])\}^{wt} - \{m_{U\text{-}F}([Gdn]_{50\%} - [Gdn])\}^{mutant} \quad (3.9)$$

For the purpose of these studies, $\Delta\Delta G_{U\text{-}F}{}^{[GdnHCl]}$ has been determined at 4 M GdnHCl, the concentration at which the midpoint of wild-type unfolding occurs. It can be seen in Table 3.1 that standard errors associated with $\Delta\Delta G_{U\text{-}F}{}^{4M}$ values (typically in the region of $\pm$ 0.1 kcal mol^{-1}) are much lower than for those of $\Delta\Delta G_{U\text{-}F}{}^{H_2O}$. The closer the chosen concentration of GdnHCl is to the value of $[GdnHCl]_{50\%}$, the shorter the extrapolation is, and so the standard error in $\Delta G_{U\text{-}F}{}^{[GdnHCl]}$ will be lower. It should be noted that for the very destabilised hydrophobic core mutants, the error associated with $\Delta\Delta G_{U\text{-}F}{}^{4M}$ is slightly higher ($\pm$ 0.2 - 0.4 kcal mol^{-1}) than for other mutants since the midpoint of unfolding for these destabilised mutants is much lower (2 M GdnHCl) than for other mutations.

In the third case, $\Delta\Delta G_{U\text{-}F}{}^{<m_{U\text{-}F}>}$, the change in free energy of unfolding between wild-type and mutant protein at a mean value of $[GdnHCl]_{50\%}$ and using an average value for $m_{U\text{-}F}$, can be calculated from equation 3.10:

$$\Delta\Delta G_{U\text{-}F}^{<m_{U\text{-}F}>} = <m_{U\text{-}F}> ([GdnHCl]_{50\%}^{wild\text{-}type} - [GdnHCl]_{50\%}^{mutant}) \quad (3.10)$$

where $([GdnHCl]_{50\%}^{wild\text{-}type} - [GdnHCl]_{50\%}^{mutant})$ is the difference in the midpoint of unfolding between wild-type and mutant proteins and $<m_{U\text{-}F}>$ is an average value of $m_{U\text{-}F}$ determined from multiple experiments on both wild-type and mutant protein.

The midpoints of unfolding obtained from repeated measurements on wild-type and different mutants of CI2 are found to be reproducible to within 0.02 M, and are insensitive to variables of the data fitting such as sloping baselines (see section 3.2(iv) below). For wild-type CI2, the average value of $[GdnHCl]_{50\%}$ obtained from four separate experiments on wild-type protein is 4.00 ± 0.01 M.

The values of $m_{U\text{-}F}$ obtained for mutants of CI2 do not vary greatly with respect to one another (see discussion below). An average $m_{U\text{-}F}$ value of 1.98 ± 0.03 kcal L mol^{-2} is obtained from repeated measurements on wild-type and a large number of mutants, and has a very low standard error. As a result values of $\Delta\Delta G_{U\text{-}F}^{<m_{U\text{-}F}>}$ have lower standard errors (typically $\pm$ 0.08 kcal mol^{-1}) than $\Delta\Delta G_{U\text{-}F}$ calculated using individual values of $m_{U\text{-}F}$ (see Table 3.1).

(iii) $\Delta\Delta G_{U\text{-}F}$ Calculated Using $<m_{U\text{-}F}>$ *versus* Individual Values of $m_{U\text{-}F}$:

As already stated, $m_{U\text{-}F}$ is a constant that is proportional to the increase in degree of solvent exposure that the protein experiences upon denaturation. For most mutants of CI2, $m_{U\text{-}F}$ is the same within experimental error as the value observed for wild-type. The same phenomenon has been observed for barnase, where virtually all barnase mutants have the same value of $m_{U\text{-}F}$ as the wild-type (Serrano *et al.*, 1992b), although this is not necessarily a general observation for all proteins. Values of $m_{U\text{-}F}$ for staphylococcal nuclease, for example, alter quite dramatically when this protein is mutated (Shortle *et al.*, 1990).

As stated above, $\Delta\Delta G_{U\text{-}F}^{H_2O}$ has been calculated using individual values $\overset{o\,f}{\underset{\wedge}{m}}_{U\text{-}F}$ (equation 3.8), resulting in standard errors that are higher than the errors associated

with $\Delta\Delta G_{U\text{-}F}^{<m_{U\text{-}F}>}$, the latter having been calculated using an average $<m_{U\text{-}F}>$ value (equation 3.10). $\Delta\Delta G_{U\text{-}F}^{<m_{U\text{-}F}>}$ is a very good approximation of $\Delta\Delta G_{U\text{-}F}^{H_2O}$ when there is no significant change in the value of $m_{U\text{-}F}$ upon mutation, i.e., when $m_{U\text{-}F}^{wt}$ and $m_{U\text{-}F}^{mutant}$ are the same as $<m_{U\text{-}F}>$, within error (since equation 3.8 becomes equation 3.10). This holds true for the hydrophobic core mutants in this study, as well as the helix mutants SA31, EQ33, ED33, EN33, and EQ34 and EA33EA34. However, for the remaining helix mutants (SG31, SG31EA33EA34, SA31EA33EA34, ED34, and EN34) individual values of $m_{U\text{-}F}$ are slightly higher than $m_{U\text{-}F}$ calculated for wild-type, falling outside the error limits. In the case of these mutants, the value of $\Delta\Delta G_{U\text{-}F}^{<m_{U\text{-}F}>}$ is not such a good approximation of $\Delta\Delta G_{U\text{-}F}^{H_2O}$ and it may be necessary to look at $\Delta\Delta G_{U\text{-}F}^{H_2O}$ directly. Unfortunately, the error associated with $\Delta\Delta G_{U\text{-}F}^{H_2O}$ for these mutants is often larger than the value $\Delta\Delta G_{U\text{-}F}^{H_2O}$ itself, rendering the data extrapolated to water very difficult to interpret. In such cases, it is possible to look at $\Delta\Delta G_{U\text{-}F}^{4M}$ instead.

Table 3.1 shows that for all mutants values for $\Delta\Delta G_{U\text{-}F}^{4M}$ are the same within error and $\Delta\Delta G_{U\text{-}F}^{<m_{U\text{-}F}>}$. For all of the hydrophobic core mutants, and for many of the helix mutants, $\Delta\Delta G_{U\text{-}F}^{H_2O}$ is also the same as $\Delta\Delta G_{U\text{-}F}^{<m_{U\text{-}F}>}$ within error. However, the error associated with $\Delta\Delta G_{U\text{-}F}^{H_2O}$ is so high as to render independent interpretation of these data very difficult without additional information, such as $\Delta\Delta G_{U\text{-}F}^{<m_{U\text{-}F}>}$.

(iv) Sloping *versus* Constant Baselines and Data Accuracy:

It was assumed in previous studies of CI2 wild-type that the fluorescence of the native and denatured forms of the protein are independent of GdnHCl concentration and so the baselines of the two-state transition were taken to be constant (Jackson and Fersht, 1991a). However, studies on mutants of CI2 suggest that this is not always the case. There appear to be small slopes associated with regions of the initial and final baselines in plots of the change in fluorescence upon titration with GdnHCl where the protein is assumed to be either fully folded or fully unfolded. At low concentrations of GdnHCl, the slope is very small, while at high concentrations, the slope is slightly

greater. The slopes of the baselines also seem to vary to a small degree amongst different mutants. However, data fitted to either equation 3.3 (with constant baselines) or equation 3.7 (with sloping baselines) give the same values of $[GdnHCl]_{50\%}$ (within error) which are then used to calculate $\Delta\Delta G_{U\text{-}F}$. All values calculated here assume sloping baselines, the data having been fitted to either equations 3.4 or 3.7.

3.3 Results from GdnHCl-Induced Equilibrium Denaturation Studies

GdnHCl-induced equilibrium denaturation experiments performed on wild-type and mutants of CI2 indicate that all sets of data adhere to a two-state model for protein folding under equilibrium conditions (see Figs. 3.1a - p). All mutants destabilise the protein with respect to wild-type. GdnHCl-induced denaturation experiments have been performed previously on wild-type CI2 (Jackson & Fersht, 1991a), and the results obtained here are in agreement with earlier findings.

The values of $m_{U\text{-}F}$, $[GdnHCl]_{50\%}$, $\Delta G_{U\text{-}F}^{H_2O}$, $\Delta\Delta G_{U\text{-}F}^{H_2O}$, $\Delta\Delta G_{U\text{-}F}^{<m_{U\text{-}F}>}$ and $\Delta\Delta G_{U\text{-}F}^{4M\ GdnHCl}$ calculated for wild-type and mutants of CI2 are presented in Table 3.1 with the standard error of the mean, α, for each value (see Appendix I). Assuming that measurements are spread according to a standard distribution about the measured value, $\bar{x}$, the probability of the true mean being between $\bar{x} \pm \alpha$ is 0.683, between $\bar{x} \pm 2\alpha$ is 0.954, and between $\bar{x} \pm 3\alpha$ is 0.997.

3.4 The Effect of GdnHCl on Mutations that Alter Protein Charge

GdnHCl may have the effect of partially screening the charge of solvent accessible side chains and, in so doing, obscure the true effect that a charge mutation might have on protein stability. For the mutations Glu to Gln and Glu to Asn (at solvent-exposed positions 33 and 34) changes in CI2 stability are observed (between 0.3 and 1.1 kcal mol^{-1}), although the full effect of removing a negative charge on stability may be masked to some extent through the use of GdnHCl. Thus, for the mutations EQ33, EN33, EQ34, and EN34, it is best to look for qualitative trends in

Table 3.1: $\Delta\Delta G_{U\text{-}F}$ for Wild-type and Mutant, as determined from GdnHCl-Induced Equilibrium Denaturation Studies of CI2

Mutant	$m_{U\text{-}F}$ [a] kcal L mol^{-2}	$[GdnHCl]_{50\%}$ [a] M	$\Delta G_{U\text{-}F}^{H_2O}$ [a,b] kcal mol^{-1}	$\Delta\Delta G_{U\text{-}F}^{H_2O}$ [c] kcal mol^{-1}	$\Delta\Delta G_{U\text{-}F}^{<m_{U\text{-}F}>}$ [d] kcal mol^{-1}	$\Delta\Delta G_{U\text{-}F}^{4M\ Gdn}$ [e] kcal mol^{-1}
wild-type	1.90 ± 0.03	4.00 ± 0.01	7.61 ± 0.11			
SG31	2.10 ± 0.07	3.59 ± 0.02	7.52 ± 0.25	0.09 ± 0.27	0.82 ± 0.04	0.89 ± 0.09
SA31	1.86 ±0.08	3.54 ± 0.02	6.58 ± 0.29	1.03 ± 0.31	0.92 ± 0.11	0.86 ± 0.13
EA33EA34	2.11 ± 0.18	3.61 ± 0.03	7.61 ± 0.67	0.00 ± 0.68	0.83 ± 0.10	0.82 ± 0.17
SG31EA33EA34	2.15 ± 0.16	3.16 ± 0.03	6.79 ± 0.52	0.82 ± 0.53	1.67 ± 0.13	1.81 ± 0.15
SA31EA33EA34	2.18 ± 0.14	3.14 ± 0.02	6.86 ± 0.45	0.75 ± 0.46	1.71 ± 0.13	1.88 ± 0.14
EQ33	1.87 ± 0.10	3.85 ± 0.03	7.03 ± 0.36	0.58 ± 0.38	0.31 ± 0.05	0.30 ± 0.14
ED33	1.83 ± 0.11	3.73 ± 0.04	6.84 ± 0.43	0.77 ± 0.44	0.54 ± 0.06	0.49 ± 0.11
EN33	2.01 ± 0.10	3.64 ± 0.02	7.29 ± 0.35	0.32 ± 0.37	0.72 ± 0.06	0.73 ± 0.13
EQ34	1.84 ± 0.09	3.76 ± 0.03	6.92 ± 0.33	0.69 ± 0.35	0.48 ± 0.05	0.45 ± 0.13
ED34	2.20 ± 0.16	3.62 ± 0.03	7.98 ± 0.58	-0.37 ± 0.59	0.76 ± 0.09	0.84 ± 0.17
EN34	2.18 ± 0.11	3.45 ± 0.02	7.53 ± 0.37	0.08 ± 0.39	1.10 ± 0.07	1.21 ± 0.11
IV48	1.99 ± 0.13	3.43 ± 0.02	6.79 ± 0.43	0.82 ± 0.44	1.14 ± 0.09	1.14 ± 0.09
IA48	2.07 ± 0.19	1.99 ± 0.03	4.12 ± 0.38	3.49 ± 0.40	3.97 ± 0.08	4.14 ± 0.39
LA68	1.98 ± 0.12	2.02 ± 0.03	3.99 ± 0.26	3.62 ± 0.28	3.94 ± 0.07	3.94 ± 0.25
IA48IV76	1.83 ± 0.10	1.90 ± 0.02	3.46 ± 0.18	4.15 ± 0.20	4.17 ± 0.09	3.85 ± 0.22

[a] for wild-type CI2, from an average of four experiments, and for CI2 mutants, from an average of 2-3 experiments per mutant; [b] $\Delta G_{U\text{-}F}^{H_2O} = m_{U\text{-}F}([GdnHCl]_{50\%})$; [c] $\Delta\Delta G_{U\text{-}F}^{H_2O} = \Delta G_{U\text{-}F}^{H_2O}{}_{\text{wild-type}} - \Delta G_{U\text{-}F}^{H_2O}{}_{\text{mutant}}$; [d] $\Delta\Delta G_{U\text{-}F}^{<m_{U\text{-}F}>} = <m_{U\text{-}F}>\Delta[GdnHCl]_{50\%}$, where $<m_{U\text{-}F}> = 1.98 \pm 0.03$ is the average $m_{U\text{-}F}$ value obtained from a combination of 4 wild-type experiments and 27 experiments using the 15 different CI2 mutants listed above;
[e] $\Delta\Delta G_{U\text{-}F}^{4M\ Gdn} = \{m_{U\text{-}F}([GdnHCl]_{50\%} - 4\ M\ GdnHCl)\}_{\text{wild-type}} - \{m_{U\text{-}F}([GdnHCl]_{50\%} - 4\ M\ GdnHCl)\}_{\text{mutant}}$

changes of protein stability, rather than an explicit quantification of free energy values, in order to avoid over interpretation of the data (see Chapter 4).

In order to measure the effects of helix and hydrophobic core mutations on CI2 stability in the absence of GdnHCl, calorimetric experiments have also been performed (see below). However, as will be discussed later (section 3.6) side chains at both positions 33 and 34 are protonated under the conditions at which the calorimetry is performed, and so no light is shed as to how much charge mutations affect CI2 stability. As will be discussed in Chapter 7 (Future Work), it may be possible to use circular-dichroism to monitor thermal denaturation to measure the change in free energy of unfolding for charge mutations in the absence of GdnHCl.

3.5 Analysis of Differential Scanning Microcalorimetry Data

(i) Parameters Obtained from Thermal Denaturation Studies:

The thermodynamics of CI2 unfolding have been studied by thermal denaturation as measured by differential scanning microcalorimetry (DSC). From thermal denaturation curves (see Fig. 3.2) it is possible to calculate the temperature at which the transition of unfolding occurs (T_m), the molar excess heat capacity at T_m ($C_p(T_m)$), and the change in heat capacity on denaturation (ΔC_p). From the area underneath the transition curve a value for $\Delta H_{cal}(T_m)$ can be calculated, where T_m is the temperature at which the heat capacity is at a maximum. $\Delta C_p(T_m)$ is the difference in heat capacity between the native and denatured states and $C_p(T_m)$ is the difference in heat capacity between the maximum and the midpoint of the native and denatured heat capacities, both at T_m. Use of equation 3.11 allows the van't Hoff enthalpy at T_m ($\Delta H_{vH}(T_m)$) to be calculated from the calorimetric transition curves (Privalov, 1986; Shrake & Ross, 1990); Kitamura & Sturtevant, 1989):

$$\Delta H_{vH}(T_m) = 4RT_m^2 \frac{C_p(T_m)}{\Delta H_{cal}(T_m)} \qquad (3.11)$$

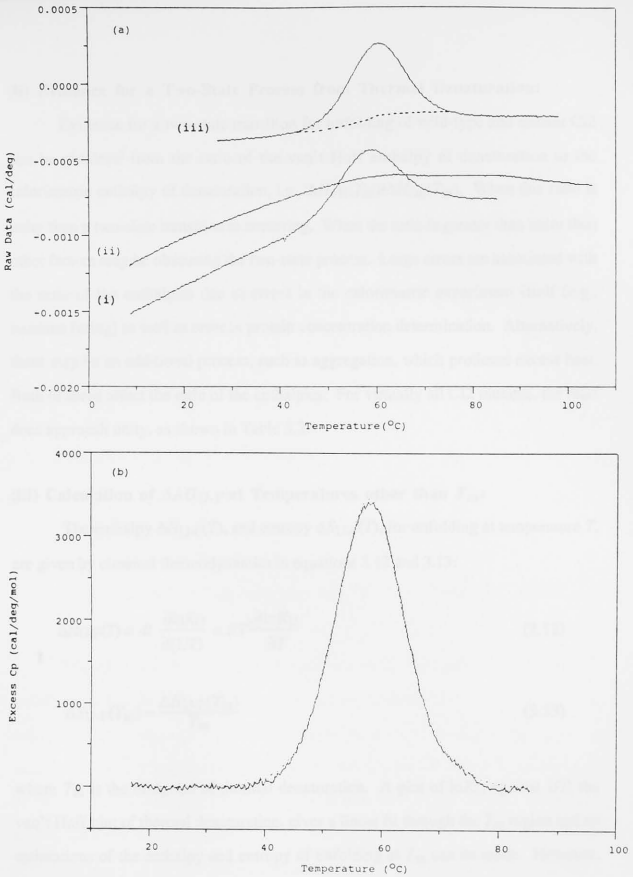


Fig. 3.2a - b: Typical DCS thermograms for the denaturation of CI2 mutant SA31 at a protein concentration of 85 μM, 10 mM glycine-Cl, pH 3.0. (a) Raw data showing I) CI2 and ii) control buffer scans with iii) CI2 data corrected for buffer baseline and indicating the interpolative baseline (dotted line) used to eliminate the denaturation heat capacity changes. (b) Normalised data showing deconvolution according to a non-two state transition model with the best fit to the data (dotted line). The T_m is 57.8 °C, calorimetric enthalpy of 55.2 kcal mol^{-1} and a van't Hoff enthalpy of 55.5 kcal mol^{-1} ($\Delta H_{vH}(T_m)/\Delta H_{cal}(T_m)$ = 1.01).

(ii) Evidence for a Two-State Process from Thermal Denaturation:

Evidence for a two-state transition for unfolding of wild-type and mutant CI2 can be obtained from the ratio of the van't Hoff enthalpy of denaturation to the calorimetric enthalpy of denaturation, i.e., $\Delta H_{vH}(T_m)/\Delta H_{cal}(T_m)$. When this ratio is unity then a two-state transition is occurring. When the ratio is greater than unity then other factors may be obscuring the two-state process. Large errors are associated with the ratio of the enthalpies due to errors in the calorimetric experiment itself (e.g., baseline fitting) as well as error in protein concentration determination. Alternatively, there may be an additional process, such as aggregation, which produces excess heat. Both of these affect the ratio of the enthalpies. For virtually all CI2 mutants, the ratio does approach unity, as shown in Table 3.2.

(iii) Calculation of $\Delta\Delta G_{U\text{-}F}$ at Temperatures other than T_m:

The enthalpy $\Delta H_{U\text{-}F}(T)$, and entropy $\Delta S_{U\text{-}F}(T)$, for unfolding at temperature T, are given by classical thermodynamics in equations 3.12 and 3.13:

$$\Delta H_{U\text{-}F}(T) = -R\,\frac{\partial \ln K_U}{\partial (1/T)} = RT^2\frac{\partial \ln K_U}{\partial T} \qquad (3.12)$$

$$\Delta S_{U\text{-}F}(T_m) = \frac{\Delta H_{U\text{-}F}(T_m)}{T_m} \qquad (3.13)$$

where T_m is the midpoint of thermal denaturation. A plot of $\ln K_U$ against $1/T$, the van't Hoff plot of thermal denaturation, gives a linear fit through the T_m region and so estimations of the enthalpy and entropy of unfolding at T_m can be made. However, over a wider range of temperatures, the plots are expected to curve due to the variation of the enthalpy and entropy of unfolding with temperature. According to equation 3.14 (Privalov, 1979):

$$\Delta C_p = \frac{\partial \Delta H_{U\text{-}F}(T)}{(\partial T)_p} = \frac{T(\partial \Delta S_{U\text{-}F}(T))}{(\partial T)_p} \qquad (3.14)$$

where ΔC_p is the change in heat capacity at constant pressure which accompanies the unfolding of a protein. In light of this, the value of $\Delta G_{U\text{-}F}(T)$ at a given temperature cannot simply be calculated from $\Delta H_{U\text{-}F}(T)$ and $\Delta S_{U\text{-}F}(T)$ without allowing for the variation due to ΔC_p. Evidence that ΔC_p is independent of temperature between 20 °C and 80 °C (Privalov, 1979) enables calculation of other thermodynamic parameters at temperatures other than T_m (providing T is 20 - 80 °C) when ΔC_p is known. Using equations 3.15, 3.16, 3.17 and 3.18 (Baldwin, 1986):

$$\Delta H_{U\text{-}F}(T) = \Delta H_{U\text{-}F}(T_m) + \Delta C_p(T - T_m) \quad (3.15)$$

$$\Delta S_{U\text{-}F}(T) = \Delta S_{U\text{-}F}(T_m) + \Delta C_p \ln(\frac{T}{T_m}) \quad (3.16)$$

$$\Delta G_{U\text{-}F}(T) = \Delta H_{U\text{-}F}(T) - T\Delta S_{U\text{-}F}(T) \quad (3.17)$$

$$\Delta G_{U\text{-}F}(T) = \Delta H_{U\text{-}F}(T_m)\{1 - (\frac{T}{T_m})\} + \Delta C_p\{(T - T_m) - T\ln(\frac{T}{T_m})\} \quad (3.18)$$

The enthalpy of unfolding at temperature T, $\Delta H_{U\text{-}F}(T)$, the entropy of unfolding at T, $\Delta S_{U\text{-}F}(T)$, and the free energy of unfolding at T, $\Delta G_{U\text{-}F}(T)$, can be calculated. Using equation 3.14 the calorimetric value for the change in heat capacity on denaturation can be calculated from the gradient of a plot of $\Delta H_{U\text{-}F}(T_m)$ versus T_m. For wild-type CI2, the value of ΔC_p is 0.72 kcal mol^{-1} K^{-1} (Jackson *et al.*, 1993a) and is assumed to be invariant with mutation. This compares well with a value of ΔC_p = 0.79 kcal mol K^{-1} for the long-form of the protein (Jackson & Fersht, 1991a), suggesting that the random coil tail contributes very little to the change in heat capacity of the protein upon denaturation. The calorimetrically determined $\Delta\Delta G_{U\text{-}F}(T)$ is simply the difference in $\Delta G_{U\text{-}F}(T)$ between wild-type and mutant proteins, and can be calculated for any temperature between 20 - 80 °C.

3.6 Results from Differential Scanning Microcalorimetry

Thermal denaturation of wild-type and mutant CI2, as followed by differential scanning microcalorimetry (DSC), was performed at pH 3.0 and/or pH 3.5. Thermodynamic parameters determined from the calorimetric data are given in Tables 3.2 and 3.3 and show that all mutants are destabilised with respect to wild-type. It can be seen in Table 3.3 that values for the change in free energy of unfolding for mutants in the hydrophobic core and at the helix N-cap of CI2 are the same, within error, when measured by thermal or GdnHCl-induced denaturation.

The results of mutations that alter the charge at positions 33 and 34 are difficult to interpret via calorimetry, however. Negatively charged side chains at these positions are protonated at pH 3.0, the pH at which the calorimetry was performed (the pK_a values of Glu33 and Glu34 are 4.0 and 3.6, respectively). To overcome this problem attempts to perform calorimtery at pH 8.0 were made, but failed due to aggregation at high temperatures when the pH is close to the pI of CI2 (6.7). It is impossible to perform calorimetry at a higher pH because no temperature insensitive buffer is available above pH 8.0. The change in free energy measured by calorimetry (at pH 3.0) on the mutants EQ33, EN33, EQ34, and EN34 are not a true reflection of how CI2 stability is affected by removing a negative charge. Hence, these values are not included in this study.

The ratio of the van't Hoff enthalpy of denaturation ($\Delta H_{vH}(T_m)$) to the calorimetric enthalpy of denaturation ($\Delta H_{cal}(T_m)$), measured at the midpoint of thermal denaturation (T_m), is close to unity for most mutants of CI2 (see Table 3.2), providing further evidence that a two-state transition for unfolding is observed.Exceptions to this are the very destabilised mutants IA48 and IA48IV76. The T_m for these mutants is very low (325 K) and so the unfolding transition coincides with strong curving of the baseline which makes baseline assignment difficult. Errors in the baseline can produce large errors in $\Delta H_{vH}(T_m)$, where an overestimation of this enthalpy will result in a high $\Delta H_{vH}(T_m)/\Delta H_{cal}(T_m)$ ratio.

Thermal denaturation of CI2 wild-type followed by DSC has been performed previously (Jackson & Fersht, 1991a) and results obtained here confirm these earlier findings. Calorimetric measurements are estimated to give values of $\Delta G_{U\text{-}F}$ that are accurate to within 5%, with contributions to the error from both the calorimetric experiment itself, such as baseline fitting, and the determination of protein concentration. $\Delta\Delta G_{U\text{-}F}$ is determined with relatively large errors, especially in the case of mutants destabilised by less than 2 kcal mol^{-1} (see Table 3.3). For most of the helix mutants as well as for the hydrophobic core mutations IV48 and LA68, $\Delta\Delta G_{cal}^{298K}$ are the same within error as values of $\Delta\Delta G_{U\text{-}F}^{<m_{U\text{-}F}>}$ calculated from GdnHCl-induced denaturation experiments. The values of $\Delta\Delta G_{cal}^{298K}$ obtained for the triple mutants SG31EA33EA34 and SA31EA33EA34 relative to the double mutant EA33EA34 give information about how the mutations SG31 and SA31 affect the stability of CI2. With respect to the double mutant, the values of $\Delta\Delta G_{cal}^{298K}$ for SA31 and SG31 are the same within error as $\Delta\Delta G_{U\text{-}F}$ from GdnHCl-induced denaturation studies.

For the mutants IA48, IA48IV76, ED33 and ED34, a value of $\Delta\Delta G_{cal}^{298K}$ is obtained that is outside of the expected error limit. IA48 and IA48IV76 are both very destabilised mutants, and as a result of low T_m (325 K), the unfolding transition coincides with strong curving of the baseline, making the baseline assignment difficult, as already stated. However, $\Delta H_{cal}(T_m)$ is not as sensitive as $\Delta H_{vH}(T_m)$ to changes in the baseline and therefore leads to a more accurate calculation of $\Delta\Delta G_{U\text{-}F}$. Thus, for these mutants, $\Delta H_{cal}(T_m)$ is used to calculate $\Delta\Delta G_{U\text{-}F}$, and this value is used for comparison with $\Delta\Delta G_{U\text{-}F}$ obtained from GdnHCl-induced denaturation (see Table 3.3). Errors for these mutants are expected to be higher (in the order of 10% for values of $\Delta G_{U\text{-}F}$), making the calorimetrically determined values of $\Delta\Delta G_{U\text{-}F}$ for IA48 and IA48IV76 the same as those from GdnHCl-induced denaturation, within this higher error limit. For mutants ED33 and ED34, calorimetric values of $\Delta\Delta G_{cal}^{298K}$ do not coincide with those obtained from GdnHCl-induced equilibrium experiments, within the initial 5% error estimate. It is possible that the concentration of ED33 and ED34 used for calculating $\Delta\Delta G_{cal}^{298K}$ was incorrect, resulting in such a discrepancy.

Table 3.2: Results of Differential Scanning Microcalorimetry Experiments Performed on Wild-type and Mutant CI2

Mutant [a]	T_m [b] K	$\Delta H_{cal}(T_m)$ [b] cal mol^{-1}	$\Delta H_{vH}(T_m)$ [c] cal mol^{-1}	$\frac{\Delta H_{vH}(T_m)}{\Delta H_{cal}(T_m)}$	$\Delta S_{cal}(T_m)$ [f] cal mol^{-1}K^{-1}	$\Delta S_{vH}(T_m)$ [g] cal mol^{-1}K^{-1}
wild-type; pH 3.5	346.84	66955	66894	0.99	193.04	192.87
wild-type; pH 3.0	336.94	60063	61487	1.02	178.26	182.49
SG31; pH 3.5	342.21	64453	63624	0.99	188.34	185.92
SA31	331.12	52957	56853	1.07	159.93	171.70
EA33EA34	333.30	58537	59049	1.01	175.63	177.16
SG31EA33EA34	329.70	55073	54053	0.98	167.04	163.95
SA31EA33EA34	329.37	53545	56272	1.05	162.57	170.85
EQ33	329.86	59012	52367	0.89	178.90	158.76
ED33	336.51	49169	58910	1.19	146.27	175.25
EN33	334.55	58589	59526	1.02	175.13	177.93
EQ34	334.48	52473	62565	1.19	156.88	187.05
ED34	335.28	60839	57807	0.95	181.46	172.41
EN34	331.13	59901	55535	0.93	180.90	167.71
IV48; pH 3.5	342.14	62833	61598	0.99	183.65	180.04
IA48; pH 3.5	324.85	35160[d]	50663[d]	1.44[e]	108.23	155.96
LA68; pH 3.5	325.30	41854	47366	1.13	128.66	145.61
IA48IV76; pH 3.5	325.79	35232[d]	48291[d]	1.37[e]	108.14	148.23

[a] all DSC experiments performed at pH 3.0, unless otherwise stated; [b] $T_m \pm 0.2$ K; [c] 5% error estimated; [d] 10% error estimated (see text); [e] ratio of enthalpies exceeds unity, most likely due to large error in enthalpies and/or protein aggregation, both of which would obscure the 2-state process (see text);

[f] $\Delta S_{cal(T_m)} = \frac{\Delta H_{cal}(T_m)}{T_m}$; [g] $\Delta S_{vH(T_m)} = \frac{\Delta H_{vH}(T_m)}{T_m}$

Table 3.3: Differences in Free Energy of Unfolding between Wild-type and Mutant CI2 as Measured by DSC and GdnHCl-Induced Equilibrium Denaturation Experiments

Mutant [a]	$\Delta G_{cal}^{298\ K}$ [b,d] kcal mol^{-1}	$\Delta G_{vH}^{298\ K}$ [b,e] kcal mol^{-1}	$\Delta\Delta G_{cal}^{298\ K}$ [f] kcal mol^{-1}	$\Delta\Delta G_{vH}^{298\ K}$ [g] kcal mol^{-1}	$\Delta\Delta G_{U\text{-}F}^{<m_{U\text{-}F}>}$ [h] kcal mol^{-1}
wild-type; pH 3.5	6.83	6.82	0.00	0.00	0.00
wild-type; pH 3.0	5.26	5.42	0.00	0.00	0.00
SG31; pH 3.5	6.18	6.07	0.65 ± 0.46	0.75 ± 0.45	0.82 ± 0.04
SA31	4.06	4.45	1.19 ± 0.40	0.97 ± 0.35	0.92 ± 0.11
EA33EA34	4.80	4.86	0.00	0.00	0.00
SG31EA33EA34	4.16	4.06	0.64 ± 0.32[i]	0.80 ± 0.31[i]	0.84 ± 0.16[i]
SA31EA33EA34	3.99	4.25	0.81 ± 0.31[i]	0.61 ± 0.32[i]	0.88 ± 0.16[i]
ED33	3.98	5.09	1.28 ± 0.33	0.33 ± 0.38	0.54 ± 0.06
ED34	5.21	4.88	0.04 ± 0.38	0.54 ± 0.36	0.76 ± 0.09
IV48; pH 3.5	5.96	5.80	0.87 ± 0.40	1.02 ± 0.40	1.14 ± 0.09
IA48; pH 3.5	2.08[c]	3.37[c]	4.74 ± 0.40	3.45 ± 0.48	3.97 ± 0.08
LA68; pH 3.5	2.66	3.13	4.16 ± 0.36	3.69 ± 0.38	3.94 ± 0.07
IA48IV76; pH 3.5	2.13[c]	3.24[c]	4.70 ± 0.40	3.58 ± 0.50	4.17 ± 0.09

[a] all DSC experiments performed at pH 3.0, unless otherwise stated; b 5% error estimated; c 10% error estimated

[d] $\Delta G_{cal}^{298\ K} = \Delta H_{cal}^{298\ K}(1 - \frac{T}{T_m}) + \Delta Cp\{(T - T_m) - Tln(1 - \frac{T}{T_m})\}$

[e] $\Delta G_{vH}^{298\ K} = \Delta H_{vH}^{298\ K}(1 - \frac{T}{T_m}) + \Delta Cp\{(T - T_m) - Tln(1 - \frac{T}{T_m})\}$, where T = 298 K& ΔCp = 0.72 kcal mol^{-1} K^{-1}

[f] $\Delta\Delta G_{cal}^{298\ K} = (\Delta G_{cal}^{298\ K})_{wild\text{-}type} - (\Delta G_{cal}^{298K})_{mutant}$

[g] $\Delta\Delta G_{vH}^{298\ K} = (\Delta G_{vH}^{298\ K})_{wild\text{-}type} - (\Delta G_{vH}^{298\ K})_{mutan\ t}$; [h] see Table 3.1; [i] relative to EA33EA34

3.7 Results and Analysis of Kinetic Unfolding Experiments

In the case of CI2, as with almost all small, single domain proteins, unfolding is found to be a single first-order reaction with the rate constant k_u (Creighton, 1988, 1990, 1993). As the final concentration of GdnHCl increases, so k_u, is found to increase, as stated in equation 3.19:

$$\ln k_u = \ln k_u{}^{H2O} + m_{\ddagger\text{-}F}[\text{GdnHCl}] \qquad (3.19)$$

where $k_u{}^{H2O}$ is the rate constant for unfolding in water, and $m_{\ddagger\text{-}F}$ is the slope of a linear fit of the unfolding rates measured over increasing concentrations of GdnHCl (for the reaction proceeding from the native (F) to the denatured (U) state, via the transition ($\ddagger$) state). Data outside of the transition region are fitted to a single exponential rate equation. As expected, the observed first order rate constant increases exponentially with increasing [GdnHCl], between 4 - 7 M GdnHCl (or 3 - 7 M for very destabilised mutants). For the purpose of analysis and reduction of errors (see later), the data are also fitted to equation 3.20 whereby the intercept shifts from 0 M GdnHCl to 4 M GdnHCl (a concentration at which unfolding kinetic rate constants are actually measured):

$$\ln k_u = \ln k_u{}^{4M} + m_{\ddagger\text{-}F}(4\text{ M GdnHCl} - [\text{GdnHCl}]) \qquad (3.20)$$

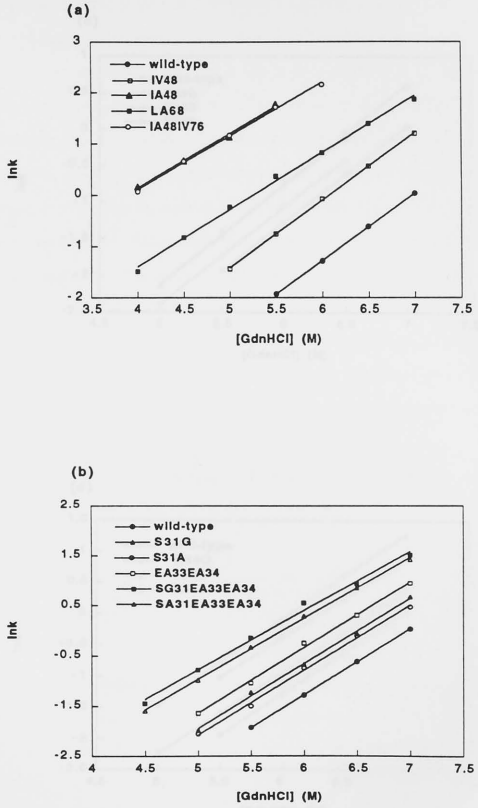
The standard errors in $\ln k_u$ when extrapolated to 4 M GdnHCl are much lower than those in water, as the extrapolation distance is shorter. Values of $\ln k_u$ (both 0 M and 4 M GdnHCl), as well as the slopes, $m_{\ddagger\text{-}F}$, for both wild-type and mutant CI2 are given in Table 3.4.

The observed rate of unfolding for mutants is faster than that for wild-type CI2 (where $k_u{}^{H2O}$ for wild-type is 1.2×10^{-4} s^{-1}). For comparative plots of $\ln k_u$ versus the final GdnHCl concentration, see figures 3.3a - d.

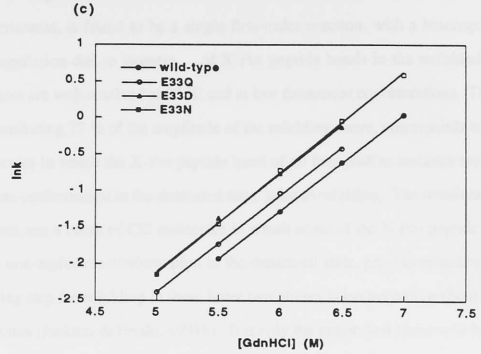
Table 3.4: Unfolding Kinetics of Wild-type and Mutant CI2

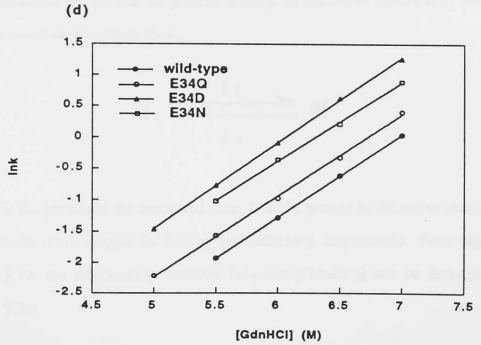
Mutant	$m_{\ddagger\text{-}F}$ [a] M^{-1}	$\ln k_u^{H_2O}$ [b]	$\ln k_u^{4M\ Gdn}$ [c]
wild-type	1.31 ± 0.01	-9.04 ± 0.07	-3.90 ± 0.02
SG31	1.30 ± 0.04	-8.42 ± 0.22	-3.24 ± 0.08
SA31	1.29 ± 0.04	-8.53 ± 0.22	-3.37 ± 0.08
EA33EA34	1.31 ± 0.04	-8.17 ± 0.23	-2.95 ± 0.08
SG31EA33EA34	1.18 ± 0.05	-6.66 ± 0.28	-1.95 ± 0.09
SA31EA33EA34	1.21 ± 0.02	-7.03 ± 0.10	-2.18 ± 0.03
EQ33	1.32 ± 0.02	-8.97 ± 0.12	-3.70 ± 0.04
ED33	1.33 ± 0.05	-8.76 ± 0.29	-3.43 ± 0.09
EN33	1.38 ± 0.02	-9.01 ± 0.13	-3.51 ± 0.05
EQ34	1.30 ± 0.03	-8.71 ± 0.16	-3.53 ± 0.05
ED34	1.37 ± 0.01	-8.28 ± 0.08	-2.82 ± 0.03
EN34	1.26 ± 0.03	-7.95 ± 0.10	-2.89 ± 0.06
IV48	1.32 ± 0.01	-8.00 ± 0.09	-2.73 ± 0.03
IA48	1.10 ± 0.04	-4.30 ± 0.18	0.10 ± 0.03
LA68	1.15 ± 0.03	-6.03 ± 0.15	-1.46 ± 0.05
IA48IV76	1.09 ± 0.03	-4.33 ± 0.15	0.04 ± 0.04

[a,b] $m_{\ddagger\text{-}F}$ is the slope and $\ln k_u^{H_2O}$ is the y-intercept from a linear fit to the GdnHCl induced unfolding kinetic data; [c] $\ln k_u^{4M\ Gdn}$ is the value of $\ln k_u$ at 4 M GdnHCl interpolated from a linear fit to GdnHCl-induced unfolding kinetic data.



Figs. 3.3a - b: Plots of $\ln k_u$ *versus* [GdnHCl] for wild-type and (a) hydrophobic core mutants IV48, IV76, LA68 and IA48IV76, and (b) helix mutants SG31, SA31, EA33EA34, SG31EA33EA34 and SA31EA33EA34.





Figs. 3.3c - d: Plots of $\ln k_u$ *versus* [GdnHCl] for wild-type and (c) helix mutants EQ33, ED33 and EN33, and (b) helix mutants EQ34, ED34, and EN34.

3.8 Results and Analysis of Kinetic Refolding Experiments

Refolding of CI2, as measured by both GdnHCl-jump and pH-jump stopped-flow experiments, is found to be a single first-order reaction, with a heterogeneous starting population due to isomers of X-Pro peptide bonds in the unfolded state. Three phases are well-resolved at 25 °C and at low denaturant concentrations. The fast phase, constituting 77 % of the amplitude of the refolding curve, corresponds to those CI2 molecules in which the X-Pro peptide bond of all four proline residues are in the native *trans*-conformation in the denatured state, prior to refolding. The remaining two, slow phases, are a result of CI2 molecules in which some of the X-Pro peptide bonds are in the non-native *cis*-conformation in the denatured state, prior to refolding. The rate-limiting step for refolding in these latter two phases is isomerisation about X-Pro peptide bonds (Jackson & Fersht, 1991b). It is only the major, fast phase which is due to protein refolding with all X-Pro peptide bonds in the *trans*-configuration in the denatured state. Only this phase is being referred to in all calculations and discussions of CI2 refolding that follow.

Assuming the process of protein folding, as measured kinetically, follows a reversible two-state transition, then:

$$\mathrm{U} \underset{k_u}{\overset{k_f}{\rightleftharpoons}} \mathrm{N}$$

where U is the protein in the denatured state, N is the protein in the native state, and k_f and k_u are the rate constants for folding and unfolding, respectively. From equations 3.2 and 3.19, the relationship between $\ln k_f$ and [GdnHCl] can be derived, as in equation 3.21:

$$\ln k_f = \ln k_f^{H_2O} - m_{\ddagger\text{-U}}[\mathrm{GdnHCl}] \qquad (3.21)$$

where $k_f^{H_2O}$ is the rate constant for refolding in water, and $m_{\ddagger\text{-}U}$ is the slope of a linear fit of the refolding rates measured over decreasing denaturant concentration (for the reaction proceeding from the denatured (U) to native (F) state, via the transition ($\ddagger$) state). Data are fitted to a double exponential rate equation. The rate constant for refolding decreases exponentially with increasing [GdnHCl] from 0 - 3 M (0 - 1.5 M for very destabilised mutants). Values of $\ln k_f^{H_2O}$ and $m_{\ddagger\text{-}U}$ for both wild-type and mutant CI2 are found in Table 3.5. The observed rate of refolding for all mutants is slower than that for wild-type CI2 (where $k_f^{H_2O}$ for wild-type is 56 s^{-1}).

Three sets of refolding data are presented in Table 3.5. The first set of data arises from extrapolation of a linear fit of all refolding data back to water, i.e., the combined rates from GdnHCl-jump and pH-jump refolding kinetics. The second arises from a linear fit of GdnHCl-jump refolding rates alone, while the third from a linear fit of pH-jump refolding rates alone. For some mutants there appears to be a small discontinuity between the rates from GdnHCl-jump experiments and those from pH-jump experiments. This results in small differences between the slopes ($m_{\ddagger\text{-}U}$) and intercepts ($\ln k_f^{H_2O}$) obtained from the different fits. For the majority of mutants, the discrepancy is small (values for $m_{\ddagger\text{-}U}$ and $\ln k_f^{H_2O}$ are the same within 10 % with the different fits). The discrepancy is slightly greater for the very destabilised mutants. Values obtained from a linear fit of the combined data are used for the purpose of this analysis and for calculating $\Delta\Delta G_{\ddagger\text{-}U}^{H_2O}$ (see Chapter 6).

3.9 Fitting Unfolding and Refolding Kinetic Data to a Two-State Model

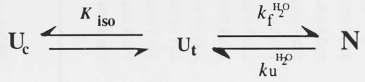
It is possible to fit the entire kinetics of unfolding and refolding to equation 3.22 (derived from equations 3.19 and 3.21), based on a two-state transition:

$$\ln k = \ln(k_f^{H_2O}\exp(-m_{\ddagger\text{-}U}[\text{GdnHCl}]) + k_u^{H_2O}\exp(m_{\ddagger\text{-}F}[\text{GdnHCl}])) \quad (3.22)$$

where k is the rate constant of unfolding or refolding at a particular concentration of denaturant, [GdnHCl]. The calculated fit of the data to equation 3.22 is shown by the

solid curve in figures 3.4a - o, the kinetics of unfolding and refolding for all mutants of CI2 adhering to this two-state model. Using the calculated fit of the data, the values of $k_f^{H_2O}$, $k_u^{H_2O}$, $m_{\ddagger\text{-}U}$, and $m_{\ddagger\text{-}F}$ can be obtained (see Table 3.6).

The most important criteria to show that a protein follows a two-state system of folding are the relationships $m_{U\text{-}F} = RT(m_{\ddagger\text{-}U} + m_{\ddagger\text{-}F})$ and $K_U = k_f/k_u$. Subsequently, the equilibrium values $\Delta G_{U\text{-}F}^{H_2O}$, $\Delta\Delta G_{U\text{-}F}^{H_2O}$, $[GdnHCl]_{50\%}$, and $m_{U\text{-}F}$ can be derived from kinetics, taking the equilibrium due to proline isomerisation into account, as follows:



where U_t is the denatured protein in which all X-Pro peptide bonds are in the *trans*-conformation, U_c is the denatured protein in which some or all of X-Pro peptide bonds are in the *cis*-conformation, and N is the native state, in which all X-Pro peptide bonds are in the *trans*-conformation. The equilibrium constant for proline isomerisation, K_{iso}, may be defined as $[U_c]/[U_t]$. The apparent equilibrium constant for unfolding calculated from the kinetic experiments, K_{app}, may be defined as $k_f^{H_2O}/k_u^{H_2O}$. From this, it follows that the equilibrium constant for unfolding, K_U, is $K_{app}/(1 + K_{iso})$. The relative amplitudes of the slow and fast refolding phases are 23 % and 77 %, respectively. Hence, $K_{iso} = 0.299$ and $K_{app} = 1.299K_U$.

Values of $\Delta G_{U\text{-}F}^{H_2O}$, $\Delta\Delta G_{U\text{-}F}^{H_2O}$, $[GdnHCl]_{50\%}$, and $m_{U\text{-}F}$ derived from a two-state fit of the unfolding plus the combined GdnHCl/pH-jump refolding kinetics are given in Table 3.6. For all single mutants, these values are the same, within error, as those obtained directly from the GdnHCl-induced equilibrium denaturation experiments. For the double and triple mutants, $\Delta G_{U\text{-}F}^{H_2O}$, $[GdnHCl]_{50\%}$, and $m_{U\text{-}F}$

Table 3.5: Refolding Kinetics of Wild-type and Mutant CI2; Data Fitted in Three Different Ways

	Combined GdnHCl and pH-jump data		GdnHCl-jump data only		pH-jump data only	
Mutant	$m_{\ddagger\text{-}U}$ [a] M^{-1}	$\ln k_f^{H_2O}$ [a]	$m_{\ddagger\text{-}U}$ [b] M^{-1}	$\ln k_f^{H_2O}$ [b]	$m_{\ddagger\text{-}U}$ [c] M^{-1}	$\ln k_f^{H_2O}$ [c]
wild-type	-1.90 ± 0.02	4.02 ± 0.02	-1.87 ± 0.03	3.94 ± 0.05	-1.89 ± 0.10	4.03 ± 0.04
SG31	-2.01 ± 0.07	3.46 ± 0.07	-1.73 ±0.03	2.93 ± 0.05	-2.22 ± 0.09	3.59 ± 0.03
SA31	-1.98 ± 0.05	3.25 ± 0.05	-1.79 ± 0.03	2.91 ± 0.05	-2.35 ± 0.09	3.40 ± 0.03
EA33EA34	-1.72 ± 0.02	3.06 ± 0.03	-1.66 ± 0.02	2.94 ± 0.03	-1.93 ± 0.09	3.13 ± 0.03
SG31EA33EA34	-1.77 ± 0.05	2.67 ± 0.04	-1.57 ± 0.06	2.37 ± 0.09	-1.88 ± 0.07	2.73 ± 0.03
SA31EA33EA34	-1.94 ± 0.04	2.78 ± 0.04	-1.80 ± 0.02	2.57 ± 0.03	-2.40 ± 0.02	2.90 ± 0.01
EQ33	-1.85 ± 0.03	3.30 ± 0.04	-1.73 ± 0.01	3.03 ± 0.02	-2.14 ± 0.07	3.43 ± 0.03
ED33	-1.95 ± 0.05	3.54 ± 0.07	-1.77 ± 0.08	3.15 ± 0.14	-2.44 ± 0.08	3.75 ± 0.03
EN33	-1.84 ± 0.04	3.00 ± 0.04	-1.74 ± 0.03	2.83 ± 0.05	-2.28 ± 0.05	3.14 ± 0.02
EQ34	-1.86 ± 0.04	3.36 ± 0.05	-1.76 ± 0.04	3.10 ± 0.07	-2.26 ± 0.09	3.52 ± 0.03
ED34	-1.90 ± 0.05	3.53 ± 0.06	-1.75 ± 0.08	3.27 ± 0.13	-2.42 ± 0.11	3.71 ± 0.04
EN34	-1.88 ± 0.04	2.93 ± 0.04	-1.79 ± 0.01	2.77 ± 0.02	-2.42 ± 0.15	3.08 ± 0.05
IV48	-2.00 ± 0.07	3.56 ± 0.06	-1.77 ± 0.02	3.28 ± 0.04	-2.65 ± 0.07	3.75 ± 0.03
IA48	-2.28 ± 0.09	2.25 ± 0.06	-1.92 ± 0.07	1.80 ± 0.07	-2.73 ± 0.13	2.37 ± 0.05
LA68	-2.20 ± 0.09	0.46 ± 0.06	-2.01 ± 0.04	0.28 ± 0.04	-2.81 ± 0.17	0.61 ± 0.05
IA48IV76	-2.66 ± 0.10	1.98 ± 0.05	-2.02 ± 0.12	1.50 ± 0.13	-2.85 ± 0.13	2.03 ± 0.05

[a] from a linear fit of the combined [GdnHCl]-jump and pH-jump refolding kinetic data, where $m_{\ddagger\text{-}U}$ is the slope of that fit and $\ln k_f^{H_2O}$ is the y-intercept; [b] from a linear fit of [GdnHCl]-jump refolding kinetic data only, where $m_{\ddagger\text{-}U}$ is the slope of that fit and $\ln k_f^{H_2O}$ is the y-intercept; [c] from a linear fit of pH-jump refolding kinetic data only, where $m_{\ddagger\text{-}U}$ is the slope of that fit and $\ln k_f^{H_2O}$ is the y-intercept.

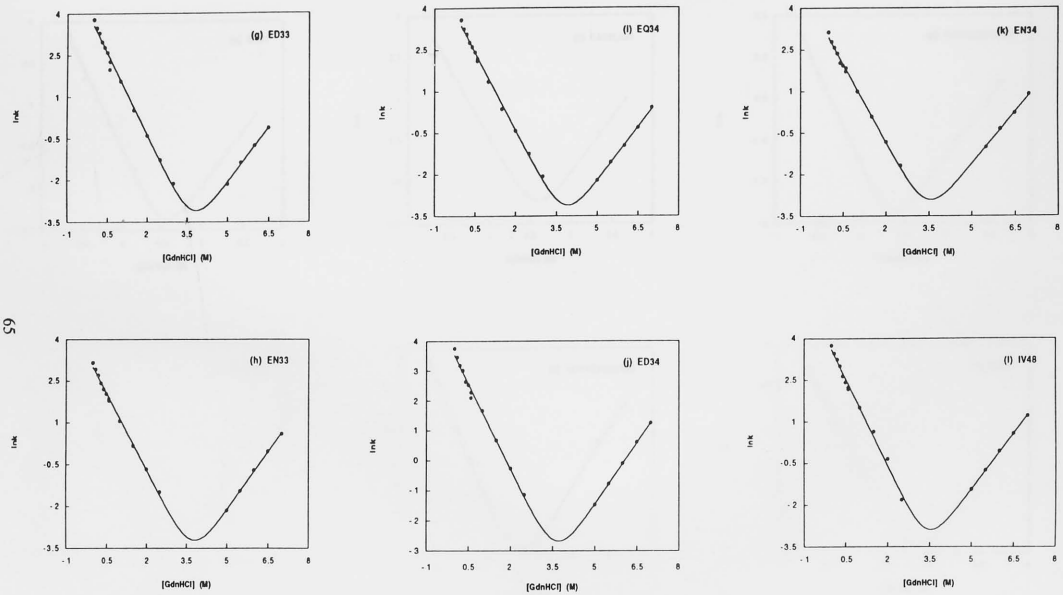
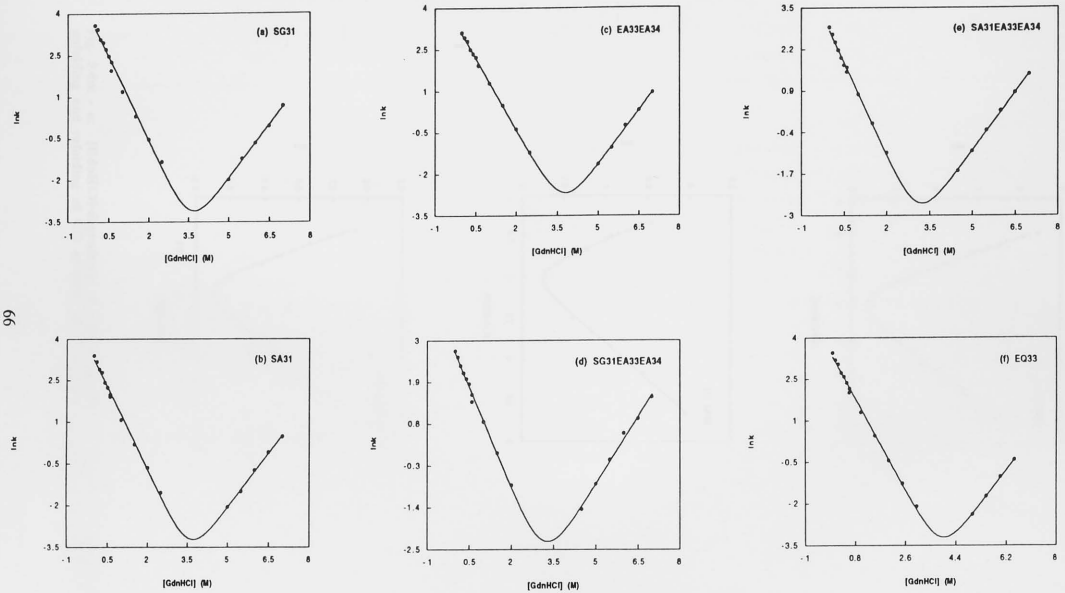
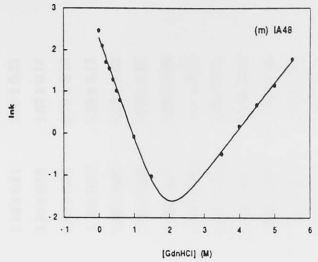
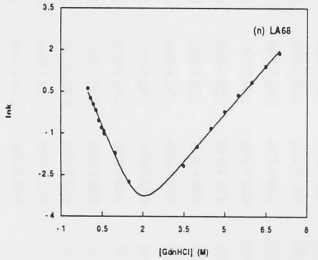


***Fig. 3.4a - f: [GdnHCl]-dependence of natural logarithm of the rate constants for unfolding and refolding of CI2 mutants (a) SG31, (b) SA31, (c) EA33EA34, (d) SG31EA33EA34, (e) SA31EA33EA34, and (f) EQ33.** * See figs. on pg. 66



* Fig. 3.4g - l: [GdnHCl]-dependence of natural logarithm of the rate constants for unfolding and refolding of CI2 mutants (g) ED33, (h) EN33, (i) EQ34, (j) ED34, (k) EN34, and (l) IV48. * see figs. on pg. 65





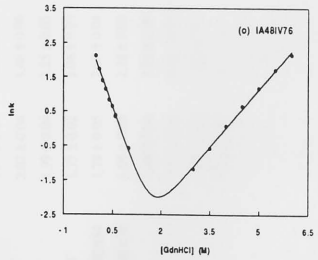


Fig. 3.4m - o: [GdnHCl]-dependence of natural logarithm of the rate constants for unfolding and refolding of CI2 mutants (m) IA48, (n) LA68, and (o) IA48IV76.

Table 3.6: Results of CI2 Unfolding Kinetics with Combined GdnHCl and pH-jump Refolding Kinetics Fitted to a Two-State Model

Mutant	$-m_{\ddagger\text{-}U}$ [a] M^{-1}	$\ln k_f^{H_2O}$ [b]	$m_{\ddagger\text{-}F}$ [c] M^{-1}	$\ln k_u^{H_2O}$ [d]	$\Delta G_{U\text{-}F}^{H_2O}$ [e] kcal mol^{-1}	$\Delta\Delta G_{U\text{-}F}$ [f] kcal mol^{-1}	$m_{U\text{-}F}$ [g] kcal L mol^{-2}	$[GdnHCl]_{50\%}$ [h] M
wild-type	1.91 ± 0.01	4.02 ± 0.02	1.32 ± 0.03	-9.19 ± 0.19	7.98 ± 0.19		1.91 ± 0.04	4.18 ± 0.08
SG31	2.02 ± 0.06	3.46 ± 0.06	1.29 ± 0.10	-8.40 ± 0.58	7.18 ± 0.59	0.80 ± 0.62	1.96 ± 0.11	3.66 ± 0.22
SA31	1.99 ± 0.04	3.25 ± 0.05	1.29 ± 0.07	-8.52 ± 0.44	7.13 ± 0.44	0.85 ± 0.48	1.94 ± 0.08	3.68 ± 0.16
EA33EA34	1.73 ± 0.02	3.06 ± 0.03	1.31 ± 0.04	-8.22 ± 0.24	6.83 ± 0.24	1.15 ± 0.30	1.80 ± 0.05	3.79 ± 0.10
SG31EA33EA34	1.78 ± 0.05	2.68 ± 0.04	1.18 ± 0.05	-6.69 ± 0.27	5.70 ± 0.27	2.28 ± 0.33	1.76 ± 0.07	3.25 ± 0.12
SA31EA33EA34	1.95 ± 0.04	2.78 ± 0.03	1.22 ± 0.03	-7.04 ± 0.20	5.97 ± 0.20	2.01 ± 0.28	1.88 ± 0.05	3.18 ± 0.08
EQ33	1.86 ± 0.03	3.31 ± 0.04	1.32 ± 0.09	-8.98 ± 0.51	7.43 ± 0.51	0.55 ± 0.55	1.88 ± 0.09	3.95 ± 0.20
ED33	1.97 ± 0.05	3.55 ± 0.06	1.31 ± 0.3	-8.66 ± 0.80	7.39 ± 0.80	0.59 ± 0.82	1.94 ± 0.15	3.80 ± 0.30
EN33	1.85 ± 0.03	3.00 ± 0.03	1.38 ± 0.05	-9.04 ± 0.32	7.28 ± 0.32	0.70 ± 0.37	1.91 ± 0.06	3.81 ± 0.12
EQ34	1.88 ± 0.03	3.36 ± 0.04	1.29 ± 0.07	-8.69 ± 0.41	7.29 ± 0.42	0.69 ± 0.46	1.88 ± 0.08	3.88 ± 0.16
ED34	1.90 ± 0.05	3.53 ± 0.05	1.37 ± 0.08	-8.29 ± 0.47	7.15 ± 0.47	0.83 ± 0.51	1.94 ± 0.09	3.69 ± 0.18
EN34	1.90 ± 0.03	2.94 ± 0.04	1.25 ± 0.08	-7.86 ± 0.47	6.55 ± 0.48	1.43 ± 0.51	1.86 ± 0.08	3.52 ± 0.18
IV48	2.10 ± 0.05	3.60 ± 0.06	1.33 ± 0.09	-8.07 ± 0.53	7.06 ± 0.53	0.92 ± 0.56	2.03 ± 0.10	3.48 ± 0.19
IA48	2.44 ± 0.08	2.29 ± 0.04	1.08 ± 0.05	-4.21 ± 0.24	4.01 ± 0.24	3.97 ± 0.31	2.09 ± 0.09	1.92 ± 0.08
LA68	2.34 ± 0.08	0.48 ± 0.05	1.14 ± 0.03	-6.03 ± 0.15	4.01 ± 0.16	3.97 ± 0.25	2.06 ± 0.09	1.94 ± 0.07
IA48IV76	2.73 ± 0.09	1.99 ± 0.03	1.12 ± 0.03	-3.52 ± 0.16	3.98 ± 0.14	4.00 ± 0.23	2.28 ± 0.09	1.75 ± 0.05

[a,b] $m_{\ddagger\text{-}U}$ and $\ln k_f^{H_2O}$ are the slope and intercept of the refolding kinetic data fitted to a 2-state model; [c,d] $m_{\ddagger\text{-}F}$ and $\ln k_u^{H_2O}$ are the slope and intercept of the unfolding kinetic data fitted to a 2-state model; [e] $\Delta G_{U\text{-}F}^{H_2O} = -RT\ln(1.299 k_f^{H_2O}/k_u^{H_2O})$; [f] $\Delta\Delta G_{U\text{-}F}^{H_2O} = (\Delta G_{U\text{-}F}^{H_2O})_{\text{wild-type}} - (\Delta G_{U\text{-}F}^{H_2O})_{\text{mutant}}$; [g] $m_{U\text{-}F} = RT(m_{\ddagger\text{-}F} + m_{\ddagger\text{-}U})$; [h] $[GdnHCl]_{50\%} = \dfrac{\ln(1.299 k_f^{H_2O}/k_u^{H_2O})}{(m_{\ddagger\text{-}U} + m_{\ddagger\text{-}F})}$

Table 3.7: Results of CI2 Unfolding Kinetics with GdnHCl-jump Refolding Kinetics Fitted to a Two-State Model of Folding

Mutant	$-m_{\ddagger\text{-}U}$ [a] M^{-1}	$\ln k_f H_2O$ [b]	$m_{\ddagger\text{-}F}$ [c] M^{-1}	$\ln k_u H_2O$ [d]	$\Delta G_{U\text{-}F}H_2O$ [e] kcal mol^{-1}	$\Delta\Delta G_{U\text{-}F}$ [f] kcal mol^{-1}	$m_{U\text{-}F}$ [g] kcal $L\ mol^{-2}$	$[GdnHCl]_{50\%}$ [h] M
wild-type	1.88 ± 0.02	3.95 ± 0.04	1.32 ± 0.03	-9.21 ± 0.17	7.95 ± 0.17		1.89 ± 0.04	4.20 ± 0.08
SG31	1.74 ± 0.03	2.94 ± 0.06	1.30 ± 0.03	-8.48 ± 0.20	6.92 ± 0.21	1.03 ± 0.27	1.80 ± 0.05	3.84 ± 0.09
SA31	1.80 ± 0.03	2.92 ± 0.06	1.30 ± 0.03	-8.58 ± 0.20	6.96 ± 0.21	0.99 ± 0.27	1.83 ± 0.05	3.80 ± 0.09
EA33EA34	1.67 ± 0.03	2.95 ± 0.05	1.32 ± 0.03	-8.24 ± 0.19	6.78 ± 0.19	1.17 ± 0.26	1.77 ± 0.04	3.84 ± 0.09
SG31EA33EA34	1.60 ± 0.09	2.39 ± 0.13	1.19 ± 0.05	-6.73 ± 0.27	5.55 ± 0.30	2.40 ± 0.34	1.65 ± 0.10	3.37 ± 0.17
SA31EA33EA34	1.82 ± 0.03	2.59 ± 0.05	1.22 ± 0.02	-7.07 ± 0.10	5.87 ± 0.11	2.08 ± 0.20	1.80 ± 0.04	3.27 ± 0.06
EQ33	1.73 ± 0.01	3.03 ± 0.02	1.34 ± 0.02	-9.12 ± 0.11	7.35 ± 0.11	0.60 ± 0.20	1.82 ± 0.02	4.04 ± 0.05
ED33	1.78 ± 0.05	3.17 ± 0.09	1.35 ± 0.09	-8.88 ± 0.51	7.29 ± 0.52	0.66 ± 0.55	1.86 ± 0.10	3.93 ± 0.21
EN33	1.75 ± 0.03	2.84 ± 0.01	1.39 ± 0.03	-8.79 ± 0.12	7.04 ± 0.12	0.91 ± 0.21	1.86 ± 0.04	3.79 ± 0.06
EQ34	1.75 ± 0.06	3.27 ± 0.10	1.39 ± 0.06	-9.11 ± 0.36	7.49 ± 0.37	0.46 ± 0.41	1.86 ± 0.08	4.02 ± 0.16
ED34	1.76 ± 0.06	3.28 ± 0.05	1.37 ± 0.06	-8.34 ± 0.1[9]	7.03 ± 0.10	0.92 ± 0.20	1.85 ± 0.08	3.79 ± 0.12
EN34	1.81 ± 0.01	2.79 ± 0.02	1.26 ± 0.02	-8.06 ± 0.14	6.58 ± 0.14	1.37 ± 0.22	1.82 ± 0.02	3.62 ± 0.06
IV48	2.09 ± 0.11	3.61 ± 0.17	1.33 ± 0.10	-8.08 ± 0.61	7.07 ± 0.63	0.88 ± 0.65	2.02 ± 0.15	3.50 ± 0.24
IA48	2.24 ± 0.21	2.11 ± 0.25	1.10 ± 0.04	-4.32 ± 0.18	3.96 ± 0.31	3.99 ± 0.35	1.98 ± 0.21	2.00 ± 0.16
LA68	2.24 ± 0.16	0.43 ± 0.17	1.15 ± 0.03	-6.04 ± 0.15	3.98 ± 0.22	3.97 ± 0.28	2.01 ± 0.17	1.98 ± 0.12
IA48IV76	2.39 ± 0.27	1.74 ± 0.13	1.12 ± 0.03	-4.49 ± 0.12	3.84 ± 0.26	4.11 ± 0.23	2.08 ± 0.27	1.84 ± 0.15

[a,b] $m_{\ddagger\text{-}U}$ and $\ln k_f H_2O$ are the slope and intercept of the refolding kinetic data fitted to a 2-state model; [c,d] $m_{\ddagger\text{-}F}$ and $\ln k_u H_2O$ are the slope and intercept of the unfolding kinetic data fitted to a 2-state model; [e] $\Delta G_{U\text{-}F}H_2O = -RT\ln(1.299k_fH_2O/k_uH_2O)$; [f] $\Delta\Delta G_{U\text{-}F}H_2O = (\Delta G_{U\text{-}F}H_2O)_{wild\text{-}type} - (\Delta G_{U\text{-}F}H_2O)_{mutant}$; [g] $m_{U\text{-}F} = RT(m_{\ddagger\text{-}F} + m_{\ddagger\text{-}U})$; [h] $[GdnHCl]_{50\%} = \dfrac{\ln(1.299k_fH_2O/k_uH_2O)}{(m_{\ddagger\text{-}U} + m_{\ddagger\text{-}F})}$

Table 3.8: Results of CI2 Unfolding Kinetics with pH-jump Refolding Kinetics Fitted to a Two-State Model of Folding

Mutant	$-m_{\ddagger\text{-}U}$ [a] M^{-1}	$\ln k_f^{H_2O}$ [b]	$m_{\ddagger\text{-}F}$ [c] M^{-1}	$\ln k_u^{H_2O}$ [d]	$\Delta G_{U\text{-}F}^{H_2O}$ [e] kcal mol^{-1}	$\Delta\Delta G_{U\text{-}F}$ [f] kcal mol^{-1}	$m_{U\text{-}F}$ [g] kcal L mol^{-2}	$[GdnHCl]_{50\%}$ [h] M
wild-type	1.89 ± 0.08	4.03 ± 0.03	1.32 ± 0.03	-9.21 ± 0.18	7.99 ± 0.18		1.90 ± 0.08	4.20 ± 0.12
SG31	2.22 ± 0.10	3.59 ± 0.04	1.30 ± 0.03	-8.44 ± 0.20	7.28 ± 0.20	0.72 ± 0.27	2.08 ± 0.10	3.49 ± 0.12
SA31	2.35 ± 0.10	3.40 ± 0.03	1.29 ± 0.03	-8.54 ± 0.20	7.22 ± 0.20	0.77 ± 0.27	2.15 ± 0.10	3.35 ± 0.11
EA33EA34	1.94 ± 0.10	3.13 ± 0.04	1.31 ± 0.03	-8.19 ± 0.21	6.86 ± 0.21	1.14 ± 0.28	1.92 ± 0.11	3.57 ± 0.13
SG31EA33EA34	1.88 ± 0.14	2.73 ± 0.05	1.18 ± 0.04	-6.69 ± 0.21	5.73 ± 0.21	2.26 ± 0.28	1.81 ± 0.14	3.16 ± 0.16
SA31EA33EA34	2.22 ± 0.09	2.87 ± 0.03	1.21 ± 0.03	-7.04 ± 0.14	6.02 ± 0.14	1.97 ± 0.23	2.03 ± 0.10	2.97 ± 0.09
EQ33	2.14 ± 0.06	3.43 ± 0.02	1.32 ± 0.03	-9.00 ± 0.18	7.52 ± 0.18	0.48 ± 0.25	2.05 ± 0.07	3.66 ± 0.09
ED33	2.44 ± 0.09	3.75 ± 0.03	1.33 ± 0.04	-8.77 ± 0.24	7.57 ± 0.24	0.42 ± 0.30	2.23 ± 0.10	3.39 ± 0.11
EN33	2.28 ± 0.06	3.14 ± 0.02	1.38 ± 0.02	-9.02 ± 0.12	7.35 ± 0.12	0.64 ± 0.22	2.16 ± 0.06	3.40 ± 0.07
EQ34	2.26 ± 0.09	3.52 ± 0.03	1.30 ± 0.03	-8.72 ± 0.18	7.40 ± 0.18	0.59 ± 0.25	2.11 ± 0.09	3.51 ± 0.10
ED34	2.42 ± 0.09	3.71 ± 0.03	1.37 ± 0.03	-8.28 ± 0.19	7.25 ± 0.18	0.74 ± 0.26	2.24 ± 0.09	3.24 ± 0.10
EN34	2.19 ± 0.15	3.04 ± 0.05	1.26 ± 0.07	-7.92 ± 0.44	6.64 ± 0.45	1.35 ± 0.48	2.04 ± 0.16	3.25 ± 0.20
IV48	2.65 ± 0.06	3.75 ± 0.02	1.32 ± 0.02	-8.00 ± 0.13	7.11 ± 0.13	0.88 ± 0.22	2.35 ± 0.07	3.03 ± 0.06
IA48	2.75 ± 0.13	2.37 ± 0.04	1.10 ± 0.04	-4.31 ± 0.19	4.11 ± 0.19	3.88 ± 0.27	2.28 ± 0.13	1.80 ± 0.08
LA68	2.73 ± 0.14	0.58 ± 0.05	1.15 ± 0.02	-6.05 ± 0.13	4.08 ± 0.14	3.91 ± 0.23	2.29 ± 0.15	1.78 ± 0.08
IA48IV76	2.88 ± 0.13	2.03 ± 0.04	1.12 ± 0.03	-4.48 ± 0.16	4.01 ± 0.13	3.99 ± 0.22	2.37 ± 0.14	1.69 ± 0.07

[a,b] $m_{\ddagger\text{-}U}$ and $\ln k_f^{H_2O}$ are the slope and intercept of the refolding kinetic data fitted to a 2-state model; [c,d] $m_{\ddagger\text{-}F}$ and $\ln k_u^{H_2O}$ are the slope and intercept of the unfolding kinetic data fitted to a 2-state model; [e] $\Delta G_{U\text{-}F}^{H_2O} = -RT\ln(1.299k_f^{H_2O}/k_u^{H_2O})$; [f] $\Delta\Delta G_{U\text{-}F}^{H_2O} = (\Delta G_{U\text{-}F}^{H_2O})_{\text{wild-type}} - (\Delta G_{U\text{-}F}^{H_2O})_{\text{mutant}}$; [g] $m_{U\text{-}F} = RT(m_{\ddagger\text{-}F} + m_{\ddagger\text{-}U})$; [h] $[GdnHCl]_{50\%} = \dfrac{\ln(1.299k_f^{H_2O}/k_u^{H_2O})}{(m_{\ddagger\text{-}U} + m_{\ddagger\text{-}F})}$

values are the same to within 10 - 20% of those obtained from equilibrium studies. Figures showing a fit of the unfolding data with i) GdnHCl-jump refolding data alone and ii) pH-jump refolding data alone are not given. However, the results of these fits can be seen in Tables 3.7 and 3.8, where $\Delta G_{U\text{-}F}{}^{H_2O}$, $\Delta\Delta G_{U\text{-}F}{}^{H_2O}$, $[GdnHCl]_{50\%}$, and $m_{U\text{-}F}$ are found to be the same as the values listed in Table 3.6 (within the region of 10%).

3.10 Discussion

Using both thermal and GdnHCl-induced equilibrium denaturation studies, it is shown that all mutants of CI2 are destabilised with respect to wild-type. For mutations that do not drastically alter the value of $m_{U\text{-}F}$, $\Delta\Delta G_{U\text{-}F}{}^{<m_{U\text{-}F}>}$, measured by GdnHCl-induced denaturation, is a very good approximation of $\Delta\Delta G_{U\text{-}F}{}^{H_2O}$ and is more precise. For most mutants of CI2, calorimetric values of $\Delta\Delta G_{U\text{-}F}{}^{H_2O}$ are the same, within error, as those values measured in the presence of GdnHCl. For the mutations that involve the removal of a negative charge, calorimetry could not be employed because the charged side chains in the wild-type protein are protonated under the conditions of the experiment. In these instances, the value of $\Delta\Delta G_{U\text{-}F}$ obtained from GdnHCl-induced denaturation is used , although the partial screening of GdnHCl on the effect of these charge mutants is noted.

CI2 is a small, monomeric, protein that folds via a sharp, two-state transition. No significantly populated intermediate state is found to exist either at equilibrium or along the kinetic folding pathway. Using thermal and GdnHCl-induced denaturation it can be seen that all mutants of CI2 observe a two-state process for equilibrium unfolding. The kinetics of unfolding and refolding of CI2 mutants, like wild-type (Jackson & Fersht, 1991b) can also be fitted to a two-state model. In addition, $m_{U\text{-}F} = RT(m_{\ddagger\text{-}U} + m_{\ddagger\text{-}F})$ and $K_U = k_f{}^{H_2O}/k_u{}^{H_2O}$, which provides strong evidence that mutants of CI2, like wild-type, follow a two-state model for protein folding at 25 °C, both kinetically and thermodynamically.

The results of the stability studies on CI2 mutants at the N-terminus of the helix are discussed in Chapter 4 of this thesis, while mutants in the hydrophobic core are discussed in Chapter 5. The kinetics of unfolding and refolding for all CI2 mutants are further analysed in Chapter 6, the data being interpreted in terms of the nature of the transition state for folding.

3.11 Conclusions

CI2 is a rare example of a protein that behaves both kinetically and thermodynamically as a two-state system, where only native and denatured states are significantly populated. Like wild-type, mutants of CI2 all adhere to this two-state model for protein folding. The two-state system is upheld at equilibrium and, for unfolding and refolding pathways, under non-equilibrium conditions also. No detectable intermediate states have been found for any of the CI2 mutants studied here, either in the equilibrium or kinetic experiments. This suggests that intermediate states, if they exist, are high in energy and only transiently populated. CI2 remains an ideal system for the purpose of studying the equilibrium and kinetics of protein folding.

Chapter 4

Importance of the Helix N-Cap Residues in CI2

4.1 The α-Helix

The α–helix is one of the most prevalent motifs in protein secondary structure. The tightly coiled polypeptide main chain (Fig. 4.1) has 3.6 residues per turn and a translation per residue of 1.50 Å, that is 5.41 Å per turn. The backbone dihedral angles, ϕ and ψ, characteristic of helices (-57, -47) are favourable for most residues (Ramachandran & Sasisekharan, 1968), the atoms of the backbone packing closely to make favourable van der Waals contacts. Backbone polar groups are automatically positioned to form hydrogen bonds with intrasegment partners when the dihedral angles of a chain segment assume helical values. This situation is unlike that of the other repetitive secondary structure, β-sheet, where backbone hydrogen bonds are satisfied by extrasegment partners from an adjacent sequence in space, but distant in sequence (Presta & Rose, 1988). The backbone carbonyl oxygen of one residue (*i*) forms a hydrogen bond with the backbone amide proton of the fourth residue along the chain (*i*+4). Typically such hydrogen bonds are 2.85 - 3.0 Å in length, as measured from the oxygen to the nitrogen atom, and are virtually straight, being nearly parallel to the axis of the helix. Not all residues in an α–helix have identical chemical properties. The first and last four residues of the α-helix differ from the rest by not being able to make the intrahelical hydrogen bonds between the backbone carbonyl groups of one turn and the amide groups of the next (Presta & Rose, 1988; Serrano & Fersht, 1989).

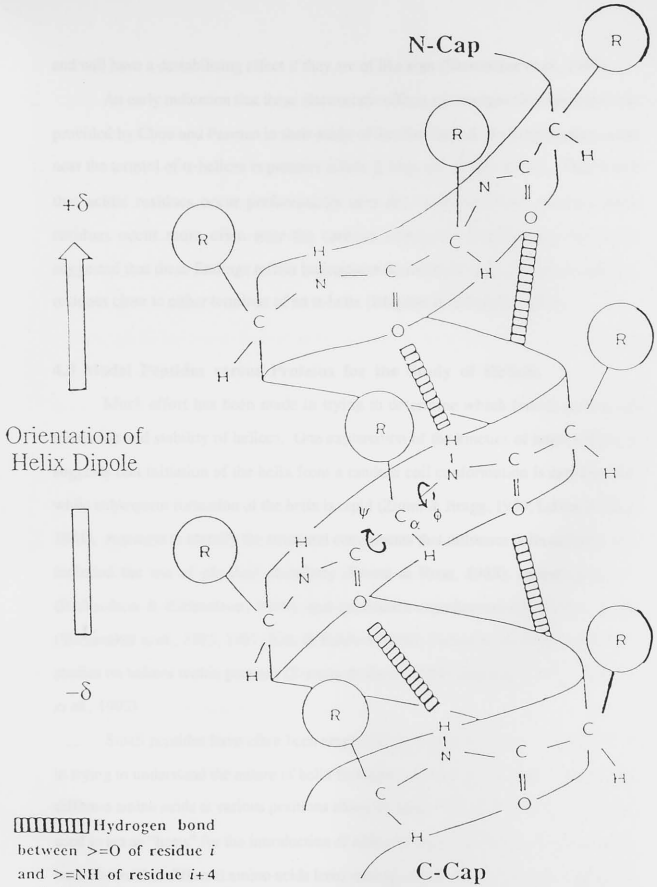
Although the screw sense of a helix can be either right-handed (clockwise) or left-handed (counter-clockwise), α-helices found in proteins are always right-handed. The side chains project away from a right-handed helix and generally do not interfere

with the backbone atoms, resulting in a sterically favourable conformation. Left-handed helices, however, are energetically less favourable due to closer contact of the side chain atoms with the helix main chain. Many α-helices are found on the surfaces of proteins, and are "amphipathic." This means that hydrophobic and hydrophilic side chains are found on opposite faces of the helix, thus rendering protein-solvent interactions favourable on the helix face consisting of polar and charged side chains, while maintaining favourable intra-protein interactions on the hydrophobic face.

In contrast to the high molecular weight helical polypeptides that were used originally to analyse helix-coil transitions and determine helical propensities (Bychkova *et al.*, 1971; Barskaya & Ptitsyn, 1971; Sueki *et al.*, 1984), single helices found in proteins are short, with an average length of about 12 resiudes (Presta & Rose, 1988; Schulz & Shirmer, 1979). Originally, isolated helices were thought to be unstable in solution, but in many cases they have now been found to be at least marginally stable in solution (Brown & Klee, 1971; Bierzynski *et al.*, 1982; Kim & Baldwin, 1984).

4.2 Interactions with the Helix Dipole

α-Helices within proteins possess a macroscopic net dipole moment (Fig. 4.1) which results from the aligned peptide bond units of the helix (Wada, 1976; Hol, 1978, 1985). The introduction of charged groups at the ends of small model helices cause changes in their melting temperature (T_m) and show qualitatively that helix stability is affected by dipole-charge interactions (Shoemaker *et al.*, 1985, 1987; Mitchinson & Baldwin, 1986). There is compelling evidence for the importance of electrostatic effects from helix dipoles on helix stability (Sali *et al.*,1988), but the precise magnitudes of these effects are unknown (Hol, 1985). The helix dipole model approximates the actual distribution of partial charges in an α–helix as an extended line dipole with its positive pole near the amino-terminus of the helix and its negative pole near the carboxy-terminus (Hol *et al.*, 1978; Sheridan *et al.*, 1982). Accordingly, the electrostatic interactions between one pole of the helix macrodipole and a nearby charged group contribute towards helix stability if the two charges are of opposite sign,



[a] adapted from Robson & Garnier, 1988

Fig. 4.1: Diagram of an α-helix showing the backbone, hydrogen bonding between residues, and orientation of the helix dipole[a]

and will have a destabilising effect if they are of like sign (Shoemaker *et al.*, 1985).

An early indication that these electrostatic effects are energetically important was provided by Chou and Fasman in their study of the distribution of charged amino acids near the termini of α-helices in proteins (Chou & Fasman, 1974a, 1974b). They found that acidic residues occur preferentially near the amino-terminus, whereas basic residues occur more often near the carboxy terminus. Blagdon and Goodman suggested that these findings reflect interactions between the helix dipole and charged residues close to either terminus of an α-helix (Blagdon & Goodman, 1975).

4.3 Model Peptides versus Proteins for the Study of Helices

Much effort has been made in trying to determine which factors govern the formation and stability of helices. One explanation of the kinetics of helix formation suggests that initiation of the helix from a random coil conformation is rate-limiting, while subsequent formation of the helix is rapid (Zimm & Bragg, 1959; Lifson & Roig, 1961). Attempts to identify the structural components that influence helix stability have included the use of physical chemistry (Presta & Rose, 1988), statistical survey (Richardson & Richardson, 1988), and qualitative experiments on model systems (Shoemaker *et al.*, 1985, 1987; Kim & Baldwin, 1984; Forood *et al.*, 1993), as well as studies on helices within proteins (Serrano & Fersht, 1989; Nicholson, 1991; Horovitz *et al.*, 1992).

Small peptides have often been employed as models for studying helices, both in trying to understand the nature of helix formation, and the contribution to stability of different amino acids at various positions along the helix. α-Helical peptides have been used to act as "hosts" for the introduction of different amino acid "guests"; such studies have shown that different amino acids have varying effects on stability (Altmann *et al.*, 1990; Padmanaban *et al.*, 1990; Padmanaban & Baldwin, 1991; Strehlow & Baldwin, 1989; Strehlow *et al.*, 1991; Marqusee & Baldwin, 1987; Marqusee *et al.*, 1989; Merutka & Stellwagen, 1990; Merutka *et al.*, 1990; O'Neil & Degrado, 1990; Lyu *et al.*, 1990, 1991; Chakrabartty *et al.*, 1991; Kemp *et al.*, 1991; Fairman *et al.*, 1991).

Attempts have been made at estabilishing a rank order for the relative helix-stabilising effects of the different amino acids and, although there are similarities amongst the studies, no single order has been unanimously agreed upon (O'Neil & Degrado, 1990; Lyu *et al.*, 1990; Chakrabartty *et al.*, 1991; Kemp *et al.*, 1991).

Several problems arise in the study of model peptides as helices. "Fraying" at the ends of helical peptides result in the existence of heterogeneous populations. The simple two-state model cannot be applied under these circumstances, making thermodynamic analysis more difficult (O'Neil & Degrado, 1990). For many single-domain proteins the two-state model is still applicable. Many investigators have studied helices directly within proteins (Matthews *et al.*, 1987; Serrano & Fersht, 1989; Ganter & Plückthun, 1990; Nicholson *et al.*, 1991; Horovitz *et al.*, 1992). Helices within proteins are more easily analysed than isolated helical peptides since protein helices are kept rigid within the overall framework of the protein's structure, obviating the problem of fraying ends (Chakrabartty *et al.*, 1991; Serrano *et al.*, 1992c). A disadvantage of studying protein helices, compared with model helical peptides in isolation, is that there may be interactons between the helix and the rest of the protein. Thus mutations designed to affect only the helix could lead to new interactions elsewhere in the protein that may obscure intrinsic effects on helix stability (Serrano *et al.*, 1992c). The possibility of such interactions must be considered when interpreting the results of studies on protein helices, especially where only limited three-dimensional structural information is available.

4.4 Helix Caps

Most of the experimental work cited above involves amino acid substitutions in the middle of the helix. The role in helix stabilisation of the first and last four residues of the α–helix, which lack intrahelical hydrogen bond partners, has also been studied (Presta & Rose, 1988; Richardson & Richardson, 1988; Serrano & Fersht, 1989; Bell *et al.*, 1991; Bruch *et al.*, 1991; Lyu *et al.*, 1992, 1993; Forood *et al.*, 1993). Research involving amino-terminal residues of model α-helical peptides (Bruch *et al.*, 1991; Lyu

et al., 1992; Forood *et al.*, 1993), as well as the site-directed mutagenesis of these residues within protein helices (Serrano & Fersht, 1989; Nicholson *et al.*, 1988; Bell *et al.*, 1991), have shown the significance of such hydrogen bonding in protein stability. It has been suggested that the location of helices in water-soluble proteins is dependent on local sequence information alone (Presta & Rose, 1988). This "helix hypothesis" states further that a necessary condition for helix formation is the presence of residues flanking the helix termini whose side chains can form hydrogen bonds with the initial four amide and final four carbonyl groups that otherwise lack intrahelical partners. Thus, there must exist a secondary structural code consisting of amino acid sequences which have the hydrogen-bonding capacity to function as helix boundaries, or "caps," and facilitate autonomous helix formation (Presta & Rose, 1988).

The N-cap (or C-cap) is defined as the first (or last) helical residue. Residues adjacent to the N-cap may be termed the N-cap+1, +2, and +3 residues (likewise, those preceeding the C-cap are called the C-cap-1, -2, and -3 residues, respectively). Richardson defines the N-cap as the first residue whose C_α lies in the axis of the α-helix (Richardson & Richardson, 1988). This includes two classes of N-cap where i) the backbone carbonyl oxygen of the N-cap residue does not hydrogen bond to the backbone amide proton of the N-cap+3 residue or ii) the backbone carbonyl oxygen (N-cap) does hydrogen bond to the backbone amide proton (N-cap+3), as appears to be the case with Ser31 at the N-cap of CI2. A statistical survey of 215 α-helices from 45 different globular protein structures (Richardson & Richardson, 1988) suggests that there are specific preferences for certain residues at the helix N-cap and C-cap which reflect the varying capacities of different side chains to fulfil the hydrogen-bonding requirements of the first and last four helical residues, respectively (Serrano *et al.*, 1989, 1992c; Richardson & Richardson, 1988; Nicholson *et al.*, 1991). The most commonly found residues at the N-cap position of helices are: Ser, Asn, Gly, Asp, and Thr. The residues Ala, Leu, Val, Ile, Trp, Arg, Gln, and Glu are found rarely. Table 4.1 shows the relative statistical preference for different amino acids at N-terminus positions along the helix, as well as in the middle and at the helix C-cap for comparison.

Table 4.1: Amino Acid Preferences At Different Positions within the Helix[§]

%	Amino Acid	N-cap	N-cap+1	N-cap+2	N-cap+3	Middle	C-cap
4.5	Pro	8[a] 0.8[b]	**25**[c] **2.6**	5 0.5	4 0.4	10[d] 0.3	7 0.7
8.7	Gly	**33** **1.8**	6 0.3	16 0.9	12 0.6	29 0.5	**72** **3.9**
6.9	Ser	**34** **2.3**	11 0.7	12 0.8	6 0.4	27 0.6	11 0.8
6.1	Thr	21 1.6	11 0.8	9 0.7	13 1.0	43 1.0	4 0.3
4.5	Asn	**34** **3.5**	7 0.7	7 0.7	7 0.7	28 0.9	15 1.6
3.6	Gln	3 0.4	5 0.7	6 0.8	12 1.5	30 1.3	7 0.9
6.0	Asp	**27** **2.1**	10 0.8	**33** **2.6**	**28** **2.2**	38 1.0	9 0.7
5.6	Glu	5 0.4	**27** **2.2**	**24** **2.0**	**40** **3.3**	30 0.8	4 0.3
6.7	Lys	10 0.7	9 0.6	14 1.0	12 0.8	49 1.1	18 1.3
3.8	Arg	3 0.4	6 0.7	7 0.9	3 0.4	33 1.3	7 0.9
2.1	His	5 1.1	3 0.7	3 0.7	3 0.7	14 1.0	6 1.3
9.0	Ala	10 0.5	24 1.2	31 1.6	19 1.0	**106** **1.8**	15 0.8
7.7	Leu	3 0.2	15 0.9	5 0.3	10 0.6	63 1.2	11 0.7
7.1	Val	1 0.1	17 1.1	9 0.6	16 1.1	56 1.2	3 0.2
5.1	Ile	2 0.2	10 0.9	8 0.7	4 0.4	41 1.2	8 0.7
3.8	Phe	2 0.2	4 0.5	7 0.9	5 0.6	34 1.3	4 0.5
3.4	Tyr	6 0.8	13 1.8	3 0.4	9 1.2	19 0.8	6 0.8
1.8	Met	3 0.8	1 0.3	4 1.0	4 1.0	18 1.5	3 0.8
1.3	Trp	1 0.3	6 2.1	5 1.7	4 1.4	14 1.5	0 0
2.4	Cys	3 0.6	4 0.8	6 1.2	3 0.6	11 0.7	2 0.4
	Total	214	214	214	194	650	212

[a] The upper entry is the observed number of occurrences out of a total number listed at the bottom of the table; [b] the lower entry is the relative preference value: the ratio of observed occurrences to the expected number based on average percentage composition (column 1); [c,d] boldface values are higher, and underscored values lower, than expectation by more than 3 standard deviations. Out of 215 helices, one amino-terminus and three carboxy-termini were omitted because their location could not be specified unambiguously.

[§] adapted from Richardson & Richardson, 1988

The following relative preferences for specific amino acids at the N-terminus of the helix are observed:

N-Cap	Ser > Gly > Ala
N-Cap+2	Asp > Glu > Ala > Asn > Gln
N-Cap+3	Glu > Asp > Ala > Gln > Asn

It should be noted that there is a marked preference for Ala in the middle of the helix, and Gly is statistically more favourable as a C-cap than as an N-cap residue.

Both kinetic and thermodymanic roles have been proposed for the N-cap residues in protein folding (Presta & Rose, 1988; Richardson & Richardson, 1988). The importance of these residues to protein stability has been shown quantitatively as well as qualitatively. Protein engineering studies on the N-cap of barnase (Serrano & Fersht, 1989), have shown that the substitution of selected residues at the N-cap can destabilise the protein by over 2 kcal mol^{-1}. Replacement of Thr6 with the residue Asp at the N-cap of a barnase helix actually stabilises the protein by 0.1 kcal mol^{-1}. The presence of a negative charge adds some 1.6 kcal mol^{-1} of stabilisation energy to the protein, measured by the stability of Asp at this position relative to Gly. This is probably due to the charge's interaction with the macroscopic electrostatic dipole of the helix. Single changes at the N-cap of T_4 lysozyme α-helices also increase the overall protein stability by as much as 1.6 kcal mol^{-1} (Nicholson *et al.*, 1988).

4.5 Structural Information About the Helix of CI2 and its N-cap

(i) The Structure of CI2 in General:

The structure of CI2 has been studied at high resolution by X-ray crystallography (McPhalen & James, 1985, 1987; Jackson *et al.*, 1993a; Harpaz *et al.*, manuscript in prepration) and by 2D NMR in solution (Clore *et al.*, 1987a; Kjaer *et al.*, 1987a, 1987b; Ludvigsen *et al.*, 1992; B. Davis, personal communication). The overall architecture of CI2 as determined by the two methods is the same (Clore *et al.*,

1987b). The main structural features of CI2 are the α-helix, an extended flexible loop, and β-sheet, both parallel and anti-parallel. The α-helix packs onto the β-sheet to form the hydrophobic core.

The α-helix of CI2 consists of 13 resiudes (31 - 43) and is not completely regular. According to the available crystal structure (McPhalen & James, 1987), some of the O···HN (*i*, *i*+4) distances are longer than usual for an α-helix (see Table 4.2). The amide nitrogens participating in long bonds may also possibly make H-bonds to the residue (*i*-3), the distorted geometry of these residues being consistent with 3_{10} helix formation. A list of the N-H···O bond lengths (distance between the nitrogen and oxygen atoms) and N-H-O angles for helix residues is given in Table 4.2 where the residue listed first in each pair is the hydrogen-bond donor. Table 4.3 lists interatomic distances between N-cap, N-cap+2 and N-cap+3 atoms and those of certain other residues in the helix. Fig. 4.2 is a diagram of the backbone structure of CI2, with side chains of N-cap, N-cap+2 and N-cap+3 residues shown.

Table 4.2: H-Bond Lengths and Angles in the α-Helix of CI2[§]

Residues[a]	NH···O length (Å)	N-H-O angle (°)
Ala35-Ser31	2.80	164
Lys36-Val32	2.90	151
Lys37-Glu33[b]	3.43	170
Val38-Glu34[b]	3.32	142
Ile39-Ala35	3.09	154
Leu40-Lys36	2.79	163
Gln41-Lys37[b]	3.44	133
Asp42-Val38	3.10	158
Lys43-Ile39	2.62	136

[a] the residue listed first in each pair is the hydrogen-bond donor; [b] distance between N and O atoms is longer than usual (where typical NH···O distance is 2.85 - 3.00 Å); residues Lys37, Val38 and Gln42 may be participating in 3_{10} helix formation.

[§] adapted from McPhalen & James, 1987

Table 4.3: Interatomic Distances between Ser31, Glu33, Glu34 and Other Helix N-terminus Residues

Atom	Neighbouring Atom	Distance (Å)	Type of Interaction
Ser31Oγ	NH (Glu34)	3.02	H-bond
Glu33Oε2	NH (Val32)	3.20	H-bond
Glu33Oε2	NH (Glu33)	2.04[a]	H-bond to itself? [a]
Glu34Oε2	NH (Ser31)	3.21	H-bond
Glu33Oε1	Cβ (Glu33)	3.78	
	NH (Glu33)	3.69	
	NH (Val32)	3.94	
	Oγ (Ser31)	3.86	
	Cβ (Ser31)	3.69	
Glu33Oε2	C (Val32)	2.78	
	Cβ (Val32)	3.06	
	Cγ (Val32)	3.30	
	O (Val32)	3.02	
	Cα (Glu33)	2.84	
	Cβ (Glu33)	2.93	
	Oγ (Ser31)	4.07	
	H_2O	3.69	
Glu34Oε1	Cβ (Glu34)	3.26	
	Oε1 (Glu33)	8.88 }	Distance between charges
	Oε2 (Glu33)	9.05 }	of Glu33 and Glu34 [b]
	Oγ (Ser31)	5.11	
Glu34Oε2	Cβ (Glu34)	3.34	
	Cβ (Ser31)	3.75	
	Cβ (Lys30)	3.93	
	Cα (Lys30)	3.90	
	Oε1 (Glu33)	7.19 }	Distance between charges
	Oε2 (Glu33)	7.61 }	of Glu33 and Glu34 [b]
	Oγ (Ser31)	3.55	

[a]The interatomic distance of 2.04 Å between the Oε2 of Glu33 and the its own backbone NH group is very short for a hydrogen bond. This would suggest that the structure of C12 is poorly defined in this region and such a contact is purely an artefact of the crystallography experiment;
[b]The great distance between the charged groups of the neighbouring Glu side chains obviates the likelihood of their experiencing any electrostatic repulsion.

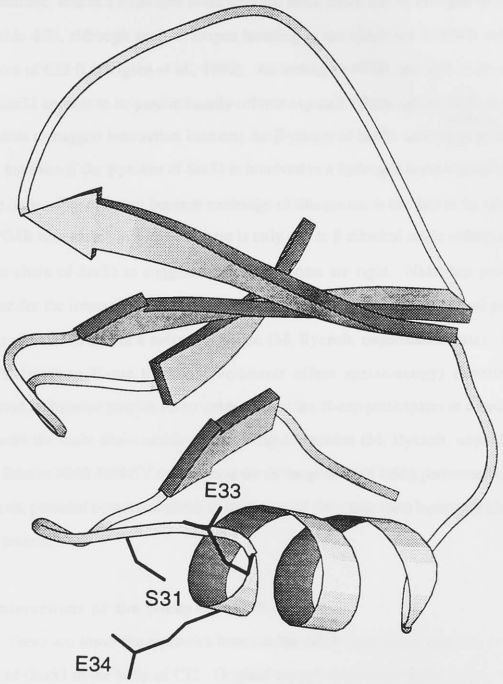


Fig. 4.2: The Backbone Structure of CI2 Showing the Side Chains of the N-cap (Ser31), N-cap+2 (Glu33), and N-cap+3 (Glu34) Residues

(ii) Interactions at the N-cap:

The crystal structure of CI2 suggests that the side chain hydroxyl of Ser31, the N-cap residue, makes a hydrogen bond with the main chain amide nitrogen of Glu34 (see Table 4.3), although such hydrogen bonding is not observed in NMR solution structures of CI2 (Ludvigsen *et al.,* 1992). According to NMR, the side chain of the N-cap Ser31 appears to be predominantly solvent exposed. There are no NOE distance constraints to suggest interaction between the β-proton of Ser31 and the amide or α-proton of Glu34, but even if the γ-proton of Ser31 is involved in a hydrogen bond with Glu34, it may be impossible to detect because exchange of this proton is too fast to be detected on an NMR timescale. In addition, there is only one α-β dihedral angle constraint for the side chain of Ser31 to suggest that these atoms are rigid. NMR has provided evidence for the interaction between N-cap and N-cap+3 residues in a helical peptide (Lyu *et al.*, 1993), and in a helix of barnase (M. Bycroft, unpublished data). NMR ROESY (rotating frame nuclear Overhauser effect spectroscopy) experiments performed on barnase provide direct evidence that the N-cap participates in a hydrogen bond wtih the main chain amide of the N-cap+3 residue (M. Bycroft, unpublished data). Similar NMR ROESY experiments are in the process of being performed on CI2 (B. Davis, personal communication) in the hopes of detecting such hydrogen bonding in this protein.

(iii) Interactions at the N-cap+2 and N-cap+3:

There are some discrepancies between the NMR and crystal structures in the region of Glu33 in the helix of CI2. Original crystal structures (McPhalen & James, 1987) predict a somewhat strained interaction between the O_ε of Glu33 and the nitrogen atom of its own backbone amide group (see Table 4.3). A distance of only 2.0 Å between these atoms (see Table 4.3) would result in an exceedingly strained conformation and is too short for a hydrogen bond measured between oxygen and nitrogen atoms. Such an interaction appears to be an artefact of the crystallography experiment (C. McPhalen, personal communication). It seems more likely that the side

chain of Glu33 shows a high degree of flexiblity in solution. This is suggested by NMR (Ludvigsen *et al.*, 1992), as only weak intraresidue NOE's are seen between the side chain and backbone atoms of Glu33 (3.3 Å NOE between the backbone amide proton and γ methylene protons of Glu33 and a 5.0 Å NOE between the α and γ methylene protons of Glu33). In addition, there appear to be no α-β dihedral angle constraints for the side chain of Glu33 in the NMR solution structure of CI2. The side chain of Glu34 appears to be fully solvated in both crystal and NMR structures. As in the case of Glu33, no α-β dihedral angle constraints exist for the side chain of Glu34, indicating that this residue is also probably flexible in solution.

(iv) Other Interactions Along the Helix:

Ala35 and Ile39 are buried residues which pack against the β-sheet to form part of the hydrophobic core. Lys36 and Lys37 appear to be fully solvated, while Val38, although not completely buried, forms part of the hydrophobic core of CI2. Val32 is partly solvated and is not sufficiently buried to be considered a "core residue." The formation of the core, and incomplete burial of Val32, are confirmed by NMR studies (Ludvigsen *et al.*, 1992). The side chains of Leu40 and Asp42 are quite flexible according to both crystal and NMR structures of CI2. However, there appears to be contact between Asp42 and both Trp24 (β-strand 1/type III turn) and Lys21 (β-strand 1). The side chain of Gln41 shows a very high degree of flexibility both in NMR and crystal structures. This is as expected for such a predominantly solvent exposed surface residue. Finally, the helix is capped at the C-terminal end by Lys43.

(v) Crystal Structures of CI2 Mutants and Hydration at the N-cap:

Mutants of CI2 have also been crystallised in an attempt to gain better structural information about this protein. Crystal stuctures of CI2 mutants SG31 and EA33EA34, SG31EA33EA34, and SA31EA33EA34 have all been solved (Harpaz, *et al.*, manuscript in preparation). However, anomalous artefacts in the 2.0 Å resolution structure of SG31 have made analysis of the helix N-cap region difficult. For example,

intermolecular hydrogen-bonds have been found to form between the N-cap hydroxyl group of one CI2 and the N-cap+3 backbone amide in the helix of a different CI2 molecule (Y. Harpaz, unpublished data). In addition, use of flexible occupancy levels in crystallographic refinement of CI2 SG31 may have resulted in B-factors that are artificially low for side chain atoms of residues in the helix (Y. Harpaz, unpublished data). Genuine flexibility in the native structure may be obscured when normally high B-factors of certain atoms are artificially lowered by the reduction in the occupancy level (e.g., from 1.0 to 0.3) of these atoms at a particular site (e.g., at positions 33 and 34; Y. Harpaz, unpublished data). Therefore, the crystallography experiment on mutant SG31 has been of little use with respect to obtaining more detailed information on the position of atoms at the N-terminal end of the helix.

Crystal structures of
The double and triple helix mutants, EA33EA34, SG31EA33EA34, and SA31EA33EA34, have been solved (to 1.74 Å resolution) without the problem of intermolecular hydrogen-bonding seen in the structure of SG31. Recent studies (Harpaz *et al.*, manuscript in preparation) have shown that crystal structures of these mutants all contain a bound water in the vicinity of the amide of Glu34 which lies progressively closer as the side chain of residue 31 is truncated. Due to the high B-factor associated with this bound water, however, it is impossible to conclude that this interaction is significant in determining the stability of the N-cap.

4.6 Description of Mutations at the α-Helix N-terminus of CI2

Mutations at the N-cap, N-cap+2, and N-cap+3 positions were designed to test the energetic bases for the statistical correlation between residue type and position in the helix presented by Richardson & Richardson (1988). The following single mutants were constructed:

N-cap	SG31; SA31
N-cap+2	EQ33; ED33; EN33
N-cap+3	EQ34; ED34; EN34

Removal of hydroxyl and methoxy groups at the Ser N-cap gives rise to the mutations Ala and Gly, respectively. Replacement of O_γ by an NH group mutates the negatively charged Glu to neutral Gln, while the additional removal of a methylene group from this side chain creates Asn. Mutation to Asp involves shortening the side chain of Glu by one CH_2 group, while maintaining a negative charge. In addition, a double and two triple mutants were made to further investigate side chain interactions in the N-cap region. In the double mutant, EA33EA34, the side chains of Glu33 and Glu34 have both been truncated to Ala in order to avoid their possible interference with the N-cap. Further mutation at the N-cap created the triple mutants SG31EA33EA34 and SA31EA33EA34.

4.7 Results of Mutations at the α-Helix N-terminus of CI2

Results from GdnHCl-induced equilibrium denaturation and calorimetric experiments performed on wild-type and mutants of CI2 have already been presented in Chapter 3 of this thesis, Tables 3.1 - 3.3. All mutants at the helix N-terminus of CI2 destabilise the protein relative to wild-type, as summarised again in Table 4.4.

At the N-cap, mutating Ser to Ala or Ser to Gly destabilises the protein by 0.92 kcal mol^{-1} and 0.82 kcal mol^{-1}, respectively. Relative to the EA33EA34 double mutant which is 0.83 kcal mol^{-1} less stable than wild-type, further mutation of Ser31 to Ala destabilises the protein by an additional 0.88 kcal mol^{-1}, while mutation to Gly destabilises CI2 by a similar additional amount, i.e., 0.84 kcal mol^{-1}.

Wild-type CI2 containing Glu at the N-cap+2 and N-cap+3 positions is more stable than mutants where Gln, Asp or Asn have been substituted at these sites. Replacement of Glu with Gln at the N-cap+2 position destabilises the protein by 0.31 kcal mol^{-1}, compared with 0.48 kcal mol^{-1} for the same mutation at the N-cap+3 site. Mutation of Glu to Asp destabilises CI2 by 0.54 kcal mol^{-1} at the N-cap+2 compared with 0.76 kcal mol^{-1} for the same mutation at the N-cap+3. Most unfavourable are Glu to Asn mutations at the N-cap+2 and N-cap+3 positions, which destabilise the protein by 0.72 kcal mol^{-1} and 1.10 kcal mol^{-1}, respectively.

The possibility of GdnHCl partially screening the full effect of charged mutants at positions 33 (N-cap+2) and 34 (N-cap+3) has been discussed in chapter 3. To avoid overinterpretation of data, a qualitative trend, rather than an explicit quantification of free energy values, is sought when discussing the effect of mutations at the N-cap+2 and N-cap+3 positions.

Table 4.4: Stability of Mutants at the Helix N-terminus

Mutant	$\Delta\Delta G_{U\text{-}F}^{<m_{U\text{-}F}>}$ [a] (kcal mol^{-1})
SG31	0.82 ± 0.04
SA31	0.92 ± 0.11
EA33EA34	0.83 ± 0.11
SG31EA33EA34	1.67 ± 0.13
SA31EA33EA34	1.71 ± 0.13
EQ33	0.31 ± 0.05
ED33	0.54 ± 0.06
EN33	0.72 ± 0.06
EQ34	0.48 ± 0.05
ED34	0.76 ± 0.09
EN34	1.10 ± 0.07

[a] See footnote to Tables 3.1 for data calculation.

4.8 Analysis and Discussion of N-cap Results

(i) Preference for Ala or Gly at the N-cap of CI2:

Mutating Ser at the N-cap of CI2 to either Ala or Gly disrupts the OHγ(N-cap) - HN(N-cap+3) H-bond, causing subsequent loss of free energy. In addition, Gly may destabilise the protein due to its preference for the unfolded state which would be more entropically favourable for this residue. There is no observable preference for Gly

compared to Ala at the CI2 N-cap as has previously been seen in barnase (Serrano & Fersht, 1989; Serrano *et al.*, 1992c) and with studies on model helices (Forood *et al.*, 1993, Lyu *et al.*, 1992), or as has been suggested by statistical survey (Richardson & Richardson, 1988) (see Table 4.5). Serrano, Fersht and coworkers (Serrano & Fersht, 1989; Serrano *et al.*, 1992c) suggest that Gly may be a better helix-capping residue than Ala because it allows for hydration at the N-cap+3 position. They proposed that hydration of the N-cap+3 backbone amide may contribute free energy similar to that lost upon mutation of the N-cap residue (Ser or Thr) and subsequent loss of the OHγ(N-cap) - HN(N-cap+3) H-bond. Ala at the N-cap, it is thought, would sterically hinder the passage of water, and not only disrupts the OHγ(N-cap) - HN(N-cap+3) H-bond but also prevents solvation that might counteract the effects of this disruption. This theory, however, does not adequately explain free energy differences among N-cap mutants in CI2. Solvation appears to be less important in stabilising the helix of CI2 than it is thought to be for helix stabilisation in barnase.

Table 4.5: Comparison of Statistical and Energetic Surveys of N-cap Stabilities

	CI2	Barnase	Statistical Survey	
Residue at the N-cap	$\Delta\Delta G_{U\text{-}F}$ [a] (kcal mol^{-1})	$\Delta\Delta G_{U\text{-}F}$ [b] (kcal mol^{-1})	Frequency at N-cap [c]	Relative Value [d]
Ser	0.0	0.0	34	2.3
Gly	0.8	1.1	33	1.8
Ala	0.9	1.9	10	0.5

[a] See Table 4.4; [b] average stabilisation energies for N-cap mutants in barnase, relative to Ser (Serrano & Fersht, 1989); [c] frequency found in statistical survey (see Table 4.1); [d] frequency at N-cap normalised for frequency of occurrence in proteins in general (see Table 4.1)

As stated in section 4.5(v) of this thesis, in the double mutant EA33EA34, a bound water molecule in the vicinity of the backbone amide of residue 34 lies progressively closer as the side chain of Ser31 is truncated from Ala to Gly (Harpaz *et al.*, manuscript in preparation). This appears to corroborate the results of Serrano & Fersht (1989) which suggest that solvation contributes to the stability of the N-cap in barnase. It is impossible to conclude, however, that solvation is significant in determining the stability of the N-cap in CI2 since the B-factor associated with the bound water is high. Attempts to identify a bound water in the vicinity of the N-cap of CI2 in the presence of Glu33 and Glu34 side chains failed, due to anomalous artefacts in the crystal structure of SG31, and the inability to crystallise the mutant SA31 (Y. Harpaz, unpublished data).

(ii) Correlating the Effect of Mutation at the N-cap on Stability with Change in Solvent Accessible Surface Area:

It has been shown that there is a good correlation in barnase between the relative stabilising effects of Ala and Gly at internal positions of helices with the total change in solvent-accessible hydrophobic surface area of the folded protein on mutation of Ala to Gly (Serrano *et al.*, 1992c; 1992d). This relationship has been extended by Serrano and coworkers (1992c, 1992d) to include the N-cap and C-cap by the addition of an extra term in hydrophilic surface area for the solvent exposure of the non-intramolecularly hydrogen-bonded main chain carbonyl, amide or protein side chain hydrogen bonding groups. This gives the correlation:

$$\Delta\Delta G = 1.68 - 0.041\Delta A_{HP} - 0.19\Delta A_{HB} \qquad (\text{kcal mol}^{-1})$$

where ΔA_{HP} is the change is solvent-accessible hydrophobic area in the folded protein on mutation of Gly to Ala, and ΔA_{HB} is the change in solvent-accessible area of hydrophilic groups that require hydrogen bonding, i.e., exposed backbone NH groups at the helix N-terminus plus surrounding side chain and main chain atoms.

In the barnase study, the preferences found at the N-cap, as determined by hydrogen bonding of side chains or solvent to the exposed backbone NH groups, are

as follows: Thr (wild-type) ~ Ser > Gly > Ala (Serrano *et al.*, 1992c, 1992d). Similar calculations performed on CI2 (based on the mutant EA33EA34; see Table 4.6), propose that Gly is marginally less destabilising than Ala at the N-cap (by 0.02 kcal mol^{-1}). This result correlates well with the data obtained from both GdnHCl-induced and thermal denaturation studies, where Gly and Ala at the N-cap of CI2 destabilise the protein by the same amount, within error.

(iii) Formation of a "Surrogate N-cap" in CI2:

Work on the N-cap of T_4 lysozyme has resulted in similar findings to those obtained for CI2 (Bell *et al.*, 1992). In T_4 lysozyme, mutating the N-cap Thr59 destabilises the protein by around 1.5 kcal mol^{-1} for both Ala and Gly mutations (*c.f.* around 1.0 kcal mol^{-1} for CI2). Although the work on T_4 lysozyme (Bell *et al.*,1992) does not corroborate the finding of Serrano & Fersht (1989; 1992c) that Gly is significantly less destabilising than Ala at the N-cap, additional factors may affect the results obtained by Bell and coworkers. In T_4 lysozyme the side chain of N-cap+2 residue Asp61 may move to act as a "surrogate N-cap" when Thr59 is replaced by either Ala or Gly, substituting a side chain (N-cap+2) to backbone (N-cap+3) H-bond for the lost side chain (N-cap) to backbone (N-cap+3) H-bond (Serrano *et al.*, 1992c). The issue of solvation at the N-cap would be mooted if the H-bonding potential of the N-cap+3 residue were satisfied by such "surrogate N-capping." In barnase, the side chain of the N-cap+2 residue (Asp8 in helix 1) is constrained by a strong salt-bridge with Arg110 (Horovitz *et al.*, 1990) and so it cannot move to cap the helix, as was found in T_4 lysozyme (the equivalent side chain of Ser28 in helix 2 is solvent exposed). This would suggest that the observed order of stability for mutants at the N-cap of barnase may be a direct result of the changes made and is not influenced by other residues in the helix, as appears to be the case in T_4 lysozyme.

In CI2, the side chains of Glu33 and Glu34 could potentially interfere with the N-cap and obscure the relative helix destabilising effects of mutants at this site. In order to test the hypothesis of Bell and coworkers (1992), the potential "surrogate side

Table 4.6: Solvent Accessible Area in Ala and Gly Mutants of CI2 N-Cap (based on EA33EA34 double mutant)

Residue	Atom	Solvent Accessible area (Å^2) [a]			
		Ala31		Gly31	
		Hydrophobic [b]	Hydrophilic	Hydrophobic	Hydrophilic
Ala/Gly31	N		-4.3		-7.8
	C_α	0.0		18.6	
	C_β	40.9			
Val32	N		-0.9		-2.2
Ala33	N		-2.0		-2.4
	C_β	60.2		65.6	
Ala34	N		-1.3		-1.9
	C_β	34.1		36.5	
Total (Å^2)		135.2	-8.5	120.7	-14.3

[a] Only those atoms in the vicinity of the N-cap which are affected by truncation of the Ser31, Glu33 and Glu34 side chains are included in these calculations. All calculations were kindly performed and communicated by L. Serrano (*c.f.* Serrano *et al.*, 1992c, 1992d).

$$\Delta\Delta G = 1.68 - 0.041\Delta A_{HP} - 0.19\Delta A_{HB} = - 0.016 \text{ kcal mol}^{-1}$$

where ΔA_{HP} is the change is solvent-accessible hydrophobic area in the folded protein on mutation of Gly to Ala (135.2 Å - 120.7 Å^2), and ΔA_{HB} is the change in solvent-accessible area of hydrophilic groups that require hydrogen bonding, i.e., exposed backbone NH groups at the helix N-terminus plus surrounding side chain and main chain atoms (-8.5 Å^2 + 14.3 Å^2).

[b] Solvent-exposed hydrophobic surface area in the folded protein having Ala instead of Ser, and the hydrophilic equivalent. These were calculated using the program of Shrake and Rupley (1973), as implemented by Miller *et al.* (1987). The Ala and Gly mutants (SA31EA33EA34 and SG31EA33EA34) were constructed by deleting the surplus side chain atoms from the coordinate file of CI2 (McPhalen & James, 1987). The exposed hydrophobic or hydrophilic surface area of the whole protein was calculated from the surface generated by the locus of the centre of a sphere of radius 1.4 Å rolling over the exposed van der Waals surface. The calculated solvent accessible hydrophobic or hydrophilic area for the mutant containing Gly at position 31 is subtracted from that calculated for the mutant containing Ala at position 31 to give the value of ΔA_{HP} or the hydrophilic equivalent, ΔA_{HB}. The changes in solvent accessible area of non-intramolecularly hydrogen-bonded groups and hydrophilic groups of the protein, on mutation of Ala to Gly, were calculated for the folded protein using the program of Miller *et al.* (1987).

chains" of CI2 were removed by replacement of Glu33 and Glu34 with Ala. Experiments on the N-cap were repeated and the destabilising effects of the double mutant (EA33EA34) and triple mutants (SG31EA33EA34, SA31EA33EA34) were measured. The relative effect of mutating the N-cap Ser31 to Ala and Gly, in the absence of Glu33 and Glu34 side chains, is observed to be around 1.0 kcal mol^{-1}. This value is the same as observed when mutating the N-cap to Ser or Gly where the side chains of Glu33 and Glu34 remain intact.

The trend of Gly being a better N-cap than Ala as observed for helices in barnase (Serrano & Fersht, 1989) is not seen in CI2. Neither is it the case that the "surrogate N-cap" effect, as observed by Bell *et al.* (1992) for T_4 lysozyme, is occurring at the N-cap of CI2. The relative helix stabilising effects of Ala and Gly at the CI2 N-cap may possibly be explained, however, in terms of the total change in solvent accessible hydrophobic surface area of the folded protein on mutation of Ala to Gly, taking into account the change in solvent accessible area of hydrophilic groups that require hydrogen bonding (*c.f.* Serrano *et al.*, 1992c, 1992d).

4.9 Discussion of N-cap+2 and N-cap+3 Results:

Mutations at the N-cap+2 are slightly less destabilising than their counterparts at the N-cap+3. The rank order of helix stabilising effects is the same for amino acid substitutions at either the N-cap+2 or N-cap+3 positions: Glu > Gln > Asp > Asn (see Tables 4.7 and 4.8).

Residues with side chains that can hydrogen bond with the main chain in the unfolded state will tend to be helix-destabilising (Ptitsyn & Finkelstein, 1983; Horovitz *et al.*, 1992). Non-charged amino acids with side chains that can form hydrogen bonds in the unfolded state are Ser, Cys, Gln and Asn. Likewise, charged residues such as Glu, Asp, Lys and Arg may also have similar contributing effects on destabilising the helix. For entropic reasons, however, residues with comparatively longer side chains are less likely to hydrogen bond with the main chain in the unfolded state. As a result, residues such as Glu and Gln will tend to be less helix-destabilising than residues with

shorter side chains such as Asp and Asn. This is reflected in the relative helix-stabilising tendencies Gln > Asn and Glu > Asp found at the N-cap+2 and N-cap+3 of CI2, as well as in previous studies on barnase (Horovitz *et al.*, 1992), model peptides (O'Neil & DeGrado, 1990; Lyu *et al.*, 1990) and by statistical survey (Richardson & Richardson, 1988).

Table 4.7: Comparison of Statistical and Energetic Surveys of N-cap+2 Stabilities

	CI2	Statistical Survey	
Residue at the N-cap+2	$\Delta\Delta G_{U\text{-}F}$ [a] (kcal mol^{-1})	Frequency at N-cap+2 [b]	Relative Value [c]
Glu	0.0	24	2.0
Gln	0.3	6	0.8
Asp	0.5	33	2.6
Asn	0.7	7	0.7

[a] See Table 4.4; [b] frequency found in statistical survey (see Table 4.1); [c] frequency at N-cap+2 normalised for frequency of occurrence in proteins in general (see Table 4.1)

The preference for Glu to Gln and for Asp to Asn found in this study may reflect the N-terminal position of these residues in CI2. The negative charge of the Glu and Asp side chains is likely to interact favourably with the partial positive charge due to the helix dipole at the N-cap. The rank order Asp > Asn is in agreement with the results of O'Neil and Degrado (1990) for replacements in the middle of the helix, while

the order Glu > Gln seen in this study does not correspond to results found elsewhere (O'Neil & DeGrado, 1990; Horovitz *et al.*, 1992). This latter order is, again, most likely due to the N-terminal position of residues 33 and 34. For mutations made in a barnase helix, the opposite trend is observed. The relative helix-stabilising tendencies Asn > Asp and Gln > Glu found in the barnase study may be due to fact that mutations are near the helix C-cap (Horovitz *et al.*, 1992). The interaction of positively charged residues at, or near, the C-cap with the partial negative charge due to the helix dipole will have a stabilising effect. Another charge effect is the relatively larger repulsion of charged side chains from a bulk non-polar α-helix than from random coil (Ptitsyn & Finkelstein, 1983; Horovitz *et al.*, 1992). As the shorter side chain of Asp brings the charge closer to the helix than in the case of Glu, this may partly explain why Asp is generally found to be more destabilising than Glu, as is seen here and in other studies (O'Neil & DeGrado, 1990; Lyu *et al.*, 1990; Horovitz *et al.*, 1992).

Table 4.8: Comparison of Statistical and Energetic Surveys of N-cap+3 Stabilities

	CI2	Statistical Survey	
Residue at the N-cap+3	$\Delta\Delta G_{U\text{-}F}$ [a] (kcal mol^{-1})	Frequency at N-cap+3 [b]	Relative Value [c]
Glu	0.0	40	3.3
Gln	0.5	12	1.5
Asp	0.8	28	2.2
Asn	1.1	7	0.7

[a] See Table 4.4; [b] frequency found in statistical survey (see Table 4.1); [c] frequency at N-cap+3 normalised for frequency of occurrence in proteins in general (see Table 4.1)

The overall rank order of helix stabilising effects of the different side chains in this study (Glu > Gln > Asp > Asn) does not reflect the overall statistical preference found by the Richardsons (1988) for these side chains. The statistical preferences for these residues at the N-cap+2 position is found to be Asp > Glu >> Asn > Gln, while at the N-cap+3 position, the order Glu > Asp >> Gln > Asn is noted (see Tables 4.1, 4.6 and 4.7). However, the individual trends of Asp > Asn and Glu > Gln, seen experimentally for CI2, have also been observed statistically (Richardson & Richardson, 1988).

4.10 Conclusions

The stability of helices, and therefore proteins, is in part, but not wholly, determined by the conformational preference of the naturally occurring amino acids. Other factors, such as amino acid positioning at, or near, the helix termini, interaction with other residues in the helix, or within the rest of the protein, as well as changes in hydrogen bonding in the folded and unfolded states, contribute to helix stability. The statistical survey of Richardson and Richardson (1988) implies that, in general, electrostatic/dipole effects are very important in determining which side chains are favourable at the N-cap. Other factors, such as side chain length, appear to affect the preference for certain amino acids at the N-cap+2, and N-cap+3 positions, for example. However, these statistical observations do not hold true for every protein. Work on CI2 does not corroborate the statistical findings, nor the results obtained at the N-cap of barnase, except in as much as the fulfilment of hydrogen bonding requirements at the N-cap is important. Where similar trends in the stability of N-cap mutants are observed between CI2 and T_4 lysozyme, for example, it is not obvious that this is the result of adherence to some common rule. Once again, the environment of the residues studied influences the result obtained. The study of different proteins, such as that performed on the helix N-cap region of CI2, stresses the fact that proteins are both complicated and individual, and that stability is dependent not only on the position of the residue being considered, but also dependent on the residues around it.

Chapter 5

The Effect of Cavity Creating Mutations in the Hydrophobic Core of CI2

5.1 The Hydrophobic Effect and its Contribution to Protein Stability

The free energy differences between the folded and unfolded states of proteins are generally quite small, typically in the range of 5 - 15 kcal mol^{-1} (Privalov, 1979; Privalov & Gill, 1988). This represents a fine balance between two large free energy terms, including a major term resulting from conformational entropy that greatly favours unfolding, and counteracting terms resulting from the hydrophobic effect and specific interactions, including ion pairs and hydrogen bonds, that favour folding (Matthews, 1987; Alber, 1989; Dill, 1990a, 1990b; Serrano *et al.*, 1992a). The reduction in free energy resulting from burial of hydrophobic residues, referred to as the hydrophobic effect, is generally considered to be the major driving force in protein folding (reviewed by Dill, 1990a). Early arguments for the importance of the hydrophobic interaction in protein folding were put forward by Linderstrom-Lang (1952) and Kauzmann (1954, 1959) when it was suggested that protein folding was driven by the aversion of non-polar residues for water. The elucidation of a great number of protein structures by crystallography and NMR supports this view. Such structural work has revealed that in most globular proteins many non-polar side chains are organised into a hydrophobic core while polar and charged side chains generally face into the solution. Upon protein folding, residues with non-polar side chains are driven from water and form a molecular interior that can shield the non-polar side chains from solvent access. Such an effect is analogous to the segregation of oil in

water, although residues in proteins are unable to partition independently as they are covalently bound to their neighbours in the protein chain (Rose *et al.*, 1985).

Much has been published in recent years concerning estimates of the net contributions of hydrophobic residues to protein stability (reviewed by Rose *et al.*, 1985; Nakai *et al.*, 1988; for experimental work see Kellis *et al.*, 1988, 1989; Alber *et al.*, 1987; Yutani *et al.*, 1987; Matsumura *et al.*, 1988; Shortle *et al.*, 1990; Sandberg & Terwilliger, 1991; Serrano *et al.*, 1992a, Fersht & Serrano, 1993). Early studies examined the binding of hydrophobic side chains of amino acids to aminoacyl-tRNA synthetases (Fersht, 1985). Such work has shown that the contribution of these hydrophobic groups to binding can be as large as 3.5 kcal mol^{-1} per methylene group (Fersht, 1985). Later work has made use of mutations at a single position to probe the contribution of hydrophobic core residues to protein stability, e.g., mutation of Ile3 in T_4 lysozyme (Matsumura, 1988), substitution by each of the 20 amino acids at position 49 (a buried residue) in the α subunit of tryptophan synthase (Yutani *et al.*, 1987), mutation of Val35 and Ile47 in bacteriophage f1 gene V protein (Sandberg & Terwilliger, 1991), and mutation of two Ile residues in barnase to both Val and Ala (Kellis *et al.*, 1989). The results of these experiments have shown that the truncation of the side chains of Ile to Val, Ala and Gly, of Leu to Ala and Gly, and of Val to Ala and Gly, etc., destabilise proteins by typically 1.0 - 2.0 kcal mol^{-1} per deleted methylene group (Fersht & Serrano, 1993). These effects are additive and so removal of several methylene groups causes a larger decrease in protein stability.

Such studies (Yutani *et al.*, 1987; Matsumura, 1988; Kellis *et al.*, 1988, 1989) have suggested that there is a correlation between the contribution of hydrophobic residues to protein stability and the hydrophobicity of these residues. Estimates of amino acid hydrophobicity are based on $\Delta G_{transfer}$, the measured free energy of transferring a side chain from water to an organic solvent (Tanford, 1962; Radzicka & Wolfenden, 1988). Measurements on the core of barnase (Kellis *et al.*, 1988, 1989), however, have concluded that the contributions to stability are twice as great as predicted by transfer free energies of model compounds (Chothia, 1976; Eisenberg &

McLachlan, 1986; Ooi *et al.*, 1987). It has become clear from this and more recent work that hydrophobicity alone does not account for the variation in free energies of mutated hydrophobic residues (Shortle *et al.*, 1990; Serrano *et al.*, 1992a). In their extensive survey of the effect of mutations on the hydrophobic core of staphylococcal nuclease, Shortle and coworkers (1990) deduced that a difference in the consequence of mutating the same residue type at two different sites in the protein must originate in the different environments surrounding the side chain at the two sites. In barnase, the contribution of a buried methyl(ene) group to protein stability is found to vary between 0.9-2.1 kcal mol^{-1}, depending upon its local environment (Serrano *et al.*, 1992a). In order to accurately predict the contribution of a residue to overall protein stability it is necessary to understand what role is played by the environment of a hydrophobic residue. This is crucial in view of how small the free energy of protein folding is, i.e., 5 - 15 kcal mol^{-1} only. Such a value is close to that of the change in free energy associated with mutating, for example, a buried Ile to Ala, the cost of which is around 5 kcal mol^{-1}.

5.2 Description of the Hydrophobic Core of CI2

The hydrophobic core of CI2 is formed at the interface between the α-helix and β-sheet, and has been studied at high resolution by X-ray crystallography (McPhalen *et al.*, 1985; McPhalen & James, 1987; Jackson *et al.*, 1993a; Harpaz *et al.*, manuscript in preparation) and in solution by 2D NMR (Clore *et al.*, 1987a; Kjaer *et al.*, 1987; Kjaer & Poulsen, 1987; Ludvigsen *et al.*, 1992). Side chain conformations in the core of CI2 are well-defined as is characteristic of densely packed residues found in the interior of proteins. The hydrophobic core of CI2 (see Fig. 5.1) consists of one aromatic (Trp24) and eleven aliphatic residues (Leu27, Ala35, Val38, Ile39, Ala46, Ile48, Val66, Leu68, Val70, Ile76, and Pro80).

Trp24, a completely buried residue, is in the first β-strand and is also involved in a type III reverse turn between Trp24-Leu27. As well as being involved in the same reverse turn as Trp24, Leu27 is involved in a type II reverse turn (Leu27-Leu30) before

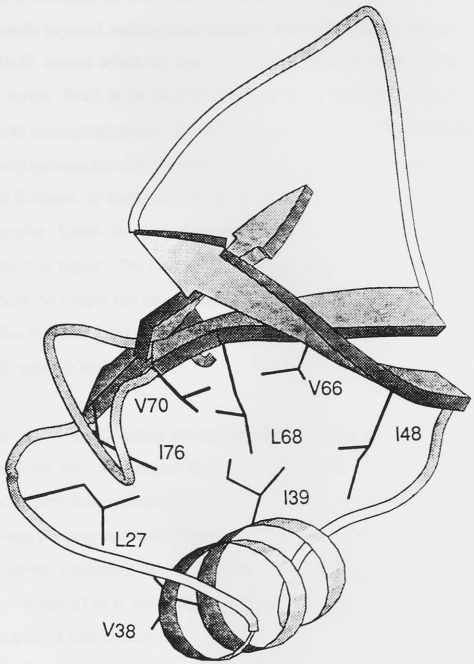


Fig. 5.1: Backbone Structure of CI2 Showing the Side Chains of Hydrophobic Core Residues Leu27, Ile39, Val38, Ile48, Val66, Leu68, Val70, and Ile76.

the start of the α-helix. Leu27 is almost entirely buried and is found on the edge of the core. Ala35 is located in the α-helix, and is completely buried. Val38, also in the α-helix, is partially exposed, making fewer contacts than other hydrophobic residues in the core. Ile39, around which the core is centred, is located in the α-helix and is completely buried. Ile48, in the second β-strand, is practically fully buried. The side chain of Ala46 is completely buried, this residue being involved in a type I reverse turn (Lys43-Ala46) between the helix and the extended loop of CI2. Val66, the first residue in the third β-strand, is completely buried, although it lies on the edge of the hydrophobic core. Leu68, also part of the third β-strand, is fully buried and lies in the middle of the core region. The final residue in this β-strand is the partially exposed Val70, its carbonyl oxygen and one of its methyl groups being partly exposed. Ile76 is part of the fourth β-strand and is entirely buried. Also in the fourth strand of β-sheet in CI2 is Pro80 which is buried.

5.3 Choice of Cavity Creating Mutations Within the Core of CI2

Mutations were constructed to test the effect of removing one or more methyl(ene) groups from particular side chains within the hydrophobic core of CI2. Mutations were designed so as to introduce "non-disruptive" deletions, and fall into one of two categories: i) mutating Ile to Val to remove a single methyl group, ii) mutating Ile to Ala or mutating Leu to Ala to remove three methyl(ene) groups. The following three single mutants were constructed:

β-strand 2	IV48; IA48
β-strand 3	LA68

Contacts are formed between the C_{δ} methyl group of Ile48 and the side chains of residues Lys36, Ile39 and Leu40 in the helix. The $C_{\gamma 1}$ and $C_{\gamma 2}$ methyl groups come into contact with Ala46, Val50, Val66 and Leu68. Mutating Ile48 to Val thus deletes contacts formed between the second β-strand and the α-helix. Truncation of Ile48 to Ala makes additional disruptions between β–strand 2 and β-strand 3. The

methyl(ene) groups of Leu68 are in contact with Val32, Ala35, Ile39, Ile48, Val50, Val66, Ile76 and Pro80. Deletion of the three methyl(ene) groups of Leu68 by mutating that side chain to Ala deletes interactions between the α-helix and β-strands 2 and 4.

In addition, the double mutant IA48IV76, which removes four methyl(ene) groups, was also analysed. In the double mutant, truncation of Ile76 to Val results in deleting contacts between β-strand 4 and both α-helix and β-strand 3, the C_δ of Ile76 being in contact with Val32, Ala35, Leu68 and Val70, while the $C_{\gamma 1}$ and $C_{\gamma 2}$ methyl groups come into contact with residues Leu27, Val28, Ala35, Leu68 and Pro80. This is in addition to the loss of interactions caused by mutating Ile48 to Ala.

A more extensive study of the core of CI2 has been carried out (Jackson *et al.*, 1993a, 1993b), of which the above four mutants are a part. Included in this study is a third type of mutation which involves the deletion of two methyl(ene) groups, i.e., the mutation of Val to Ala. For the purpose of correlating the effects of mutating core residues in CI2 with environmental parameters (see discussion later) the full complement of CI2 core mutants (listed with measured $\Delta\Delta G_{U\text{-}F}$ values in Table 5.2) will be used. The results of mutations in the core of CI2, other than those described above (i.e., IV48, IA48, LA68 and IA48IV76) have been kindly communicated by Dr S. Jackson and do not form a part of any other thesis or dissertation.

5.4 Results of Mutations in the Hydrophobic Core of CI2

Results of GdnHCl-induced denaturation and calorimetric experiments performed on wild-type and hydrophobic core mutants of CI2 have already been presented in Chpater 3 of this thesis, Tables 3.1 - 3.3. All mutants in the core of CI2 destabilise the protein relative to wild-type, as summarised again in Table 5.1.

In the second strand of the β-sheet, mutating Ile48 to Val destabilises the protein by 1.14 kcal mol^{-1}, while further mutation of Ile to Ala destabilises the protein by 3.97 kcal mol^{-1}. Mutating Leu68 to Ala, in the third β-strand, destabilises CI2 by

3.94 kcal mol^{-1}. Finally, simultaneous mutation of Ile48 to Ala and Ile76 to Val causes the protein to be destabilised by as much as 4.17 kcal mol^{-1}.

Table 5.1: Change in Free Energy of Unfolding for Hydrophobic Core Mutants, as Monitored by GdnHCl-Induced Denaturation and DSC [a]

Mutant	$\Delta\Delta G_{U\text{-}F}^{H_2O}$ kcal mol^{-1}	$\Delta\Delta G_{U\text{-}F}^{<m_{U\text{-}F}>}$ kcal mol^{-1}	$\Delta\Delta G_{cal}^{(298\ K)}$ kcal mol^{-1}
IV48	0.82 ± 0.44	1.14 ± 0.09	1.07
IA48	3.49 ± 0.40	3.97 ± 0.08	3.84
LA68	3.62 ± 0.28	3.94 ± 0.07	3.71
IA48IV76	4.15 ± 0.20	4.17 ± 0.09	4.09

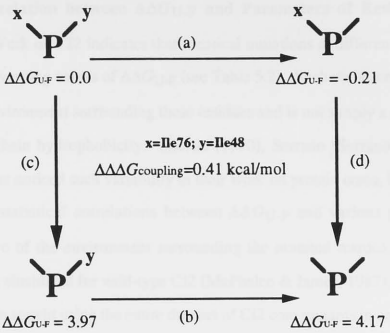
[a] Summarised from Tables 3.1-3.3; see footnotes of these tables for information on data calculation.

5.5 Analysis of Results

(i) Double Mutant Cycle Analysis of IA48IV76:

The free energy of coupling, $\Delta\Delta\Delta G_{(x,y)}$, between two residues x and *y* in a protein may be measured by employing a double mutant cycle (see Fig. 5.2). The term free energy of coupling ($\Delta\Delta\Delta G_{(x,y)}$) is used here, rather than the free energy of interaction ($\Delta\Delta\Delta G_{int}$), since the interaction between the two residues may not be direct, but instead may be indirectly propagated through the protein. The general theory of the use of double mutant cycles has been presented and discussed elsewhere (Horovitz & Fersht, 1990). From a double mutant cycle (Fig. 5.2) of the GdnHCl-induced equilibrium denaturation data for the single mutants IA48, IV76 (Jackson *et al.*, 1993a), and for the double mutant IA48IV76, the free energy of coupling can be calculated between the two residues Ile48 and Ile76 in the native state of CI2.

Fig. 5.2: Double Mutant Cycle Analysis of IA48IV76[a]



$\Delta\Delta\Delta G_{coupling}$ = (a) - (b) = (c) - (d)

= $\Delta\Delta G_{U\text{-}F}(Px,y\text{->}Px) - \Delta\Delta G_{U\text{-}F}(Py\text{->}P) = \Delta\Delta G_{U\text{-}F}(Px,y\text{->}Py) - \Delta\Delta G_{U\text{-}F}(Px\text{->}P)$

= (0.0 + 0.21) - (3.97 - 4.17)kcal/mol = (0.0 - 3.97) - (- 0.21 - 4.17) kcal/mol

= 0.41 kcal/mol coupling energy between Ile48 and Ile76

[a]adapted from Horovitz & Fersht, 1990

Both Ile48 and Ile76 have side chains that are buried in the hydrophobic core of the protein. However, these side chains are at least 8 Å apart from one another, and so are not in direct contact in the core. If the effects of the two mutations are independent, then identical changes in energy are found upon making the same mutation in wild-type and mutant enzymes, i.e., $\Delta\Delta\Delta G_{(x,y)}$ would be zero. However, a non-zero coupling energy ($\Delta\Delta\Delta G_{(x,y)}$ = 0.41 kcal mol^{-1}) seen between residues Ile48 and Ile76 in the native state of CI2 suggests that the effects of these two mutations are not independent. It would appear that mutating Ile48 to Ala in the presence of the mutation IV76 destabilises the protein to a greater extent than when Ile76 is not mutated, i.e., in the wild-type protein. This is most likely the result of small structural rearrangements in the core of the protein upon creation of a cavity when Ile76 is mutated to Val (Jackson *et al.*, 1993a).

(ii) Correlation between $\Delta\Delta G_{U\text{-}F}$ and Parameters of Residue Environment:

Work on CI2 indicates that identical mutations at different sites in the protein result in varying values of $\Delta\Delta G_{U\text{-}F}$ (see Table 5.2). Such variation must be due in part to the environment surrounding these residues and is not simply a result of differences in side chain hydrophobicity. Shortle (1990), Serrano (Serrano *et al.*, 1992a), and coworkers noticed such variability in their work on protein cores, leading these groups to seek statistical correlations between $\Delta\Delta G_{U\text{-}F}$ and various parameters that are indicative of the environment surrounding the mutated residue. Using the crystal structure elucidated for wild-type CI2 (McPhalen & James, 1987), similar correlations have been sought using the entire data set of CI2 core mutants in combination with data previously determined for barnase (Serrano *et al.*, 1992a). In addition, separate analysis of CI2 mutants alone has been made, the results of these correlations being compared to those obtained using both CI2 and barnase data. The two sets of correlations are discussed below, with several strong correlations having been found.

A correlation has been sought between the change in free energy of unfolding, $\Delta\Delta G_{U\text{-}F}$, and the difference in side chain solvent accessible surface area buried between wild-type and mutant protein (see Table 5.2 and Fig. 5.3a-b). The change in side chain solvent accessible surface area buried on folding can be estimated by assuming that the side chains are completely accessible to solvent in the unfolded protein, while the algorithm of Lee & Richards (1971) can be employed to assess which groups are buried in the folded protein. For the ten single hydrophobic core mutants of CI2, the correlation coefficient for a linear fit of the data is found to be 0.82 (see Fig. 5.3a). This compares well with results obtained from a similar correlation previously sought using 20 hydrophobic core mutants of the protein barnase (correlation coefficient of 0.83; Serrano *et al.*, 1992a). The results of CI2 mutants are combined with those of barnase in Fig. 5.3b, and it can be seen that both sets of data fall on the same line, with a strong correlation coefficient of 0.82 for the combined data. The relationship between the change in buried side chain solvent accessible surface area and $\Delta\Delta G_{U\text{-}F}$ exists for more than one protein, and appears to be general.

Table 5.2: Relationship between $\Delta\Delta G_{U\text{-}F}$ upon creating cavities in the hydrophobic core of CI2 and certain environmental parameters[§]

Mutant	Solvent accessible surface area of the wild-type side chain (Å2) [a,b]	Δsolvent accessible surface area on mutation buried by side chain (Å2) [a,c]	Number of methyl(ene)s within a 6Å radius of deleted carbons [a,d]	$\Delta\Delta G_{U\text{-}F}$<$m_{U\text{-}F}$> kcal mol^{-1}
IV48	10.8	22.7	15	1.14[e]
IA48	10.8	76.8	45	3.97[e]
LA68	0.0	73.4	49	3.94[e]
IA48IV76	10.9	99.8	95	4.17[e]
LA27	1.4	72.8	50	2.64[a]
VA38	52.0	26.4	18	0.46[a]
IV39	0.0	23.3	24	1.27[a]
VA66	0.0	50.0	39	4.88[a]
VA70	10.1	43.8	22	1.95[a]
IV76	0.1	23.0	12	-0.21[a]
IA76	0.1	73.9	40	4.25[a]

[a] Jackson *et al.* (1993a); [b] the solvent accessible area of the different target residues calculated using the algorithm of Lee & Richards (1971); this column gives the value for the exposed area on the side chain in wild-type protein; [c] calculation as described in Chapter 2 (methods and materials); [d] the number of methyl(ene) groups (excluding the C_α) at a distance of less than 6Å was determined for each methyl(ene) group deleted by mutation and then summed; [e] See Table 5.1.

[§] adapted from Serrano *et al.*, 1992.

Correlating $\Delta\Delta G_{U\text{-}F}$ with M, the number of methyl(ene) groups deleted upon mutation, gives a weak correlation for both mutants of CI2 and barnase. For the combined data a correlation coefficient of 0.67 is observed, where 30 single mutants have been analysed. Using a more refined environmental parameter, N (the number of methyl(ene) side chain groups that surround the deleted methyl(ene) groups in the mutant protein within a 6 Å radius), a much better correlation with $\Delta\Delta G_{U\text{-}F}$ is observed. For CI2 mutants alone (excluding the double mutant IA48IV76), the correlation coefficient is 0.85 (see Fig. 5.4a). Using a set of twenty mutants, a

correlation coefficient of 0.90 was found for barnase (Serrano *et al.*, 1992a). Once again, the combined CI2 and barnase data sets fall on the same line (correlation of 0.87 for 30 points; see Fig. 5.4b), suggesting that N is a good indicator of how the environment of a hydrophobic core residue can influence the extent to which its mutation affects protein stability.

In order to establish that the above correlation with N is not simply a reflection of the size of the side chain deleted, correlations were repeated using only the set of Ile to Val mutants (these tending to be the least destabilising of the hydrophobic core mutations in both CI2 and barnase; see Table 5.3). Using the combined CI2 and barnase Ile to Val mutants (omitting CI2 IV76), correlations were sought between $\Delta\Delta G_{U\text{-}F}$ and i) the difference in side chain solvent accessible surface area buried between wild-type and mutant (see Fig. 5.5a) and ii) N (see Fig. 5.5b). A correlation coefficient of 0.60 is obtained for i), which is much weaker than the value 0.82 obtained using all types of core mutation (see Figs. 5.3a-b). However, a strong correlation coefficient of 0.83 is found in the case of ii), which compares favourably with the value 0.85 - 0.87 found previously (see Figs. 5.4a-b). From this it is possible to see that the correlation between $\Delta\Delta G_{U\text{-}F}$ and N is strong regardless of the size of the group being deleted during mutation.

Using the change in free energy of transferring an amino acid side chain from water to *n*-octanol (ΔG_{tr}) (Fauchere & Pliska, 1983), corrected for changes in volume (Sharp *et al.*, 1991), Pace (1992) sought a correlation between ΔG_{tr} and $\Delta\Delta G_{U\text{-}F}$. In this analysis, an average $\Delta\Delta G_{U\text{-}F}$ was used based on mutation of side chains in the cores of barnase, staphylococcal nuclease, T_4 lysozyme and gene V protein from bacteriophage f1. Although a correlation coefficient of 0.99 was obtained, suggesting that the correlation is very strong, the standard deviation in the sample used was also high. Using $\Delta\Delta G_{U\text{-}F}$ values obtained for core mutants in CI2, a weaker correlation coefficient of 0.67 is found between $\Delta\Delta G_{U\text{-}F}$ and ΔG_{tr} (see Table 5.4 and Fig. 5.6). This is not as strong as correlations derived earlier, between $\Delta\Delta G_{U\text{-}F}$ and N.

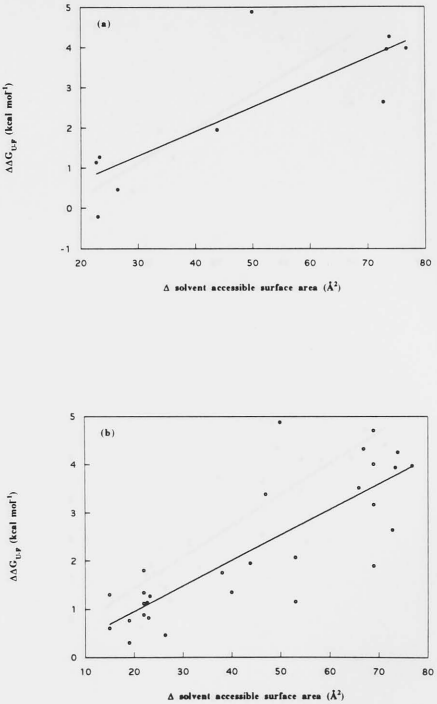


Fig. 5.3: Correlation between the difference in solvent accessible area that is buried by wild-type and mutant side chains, and the change in free energy of unfolding for mutations of hydrophobic residues in a) CI2 (slope = 0.06; R = 82) and b) CI2 and barnase (slope = 0.05; R = 0.82); filled circles (•) are CI2, open circles (o) are barnase.

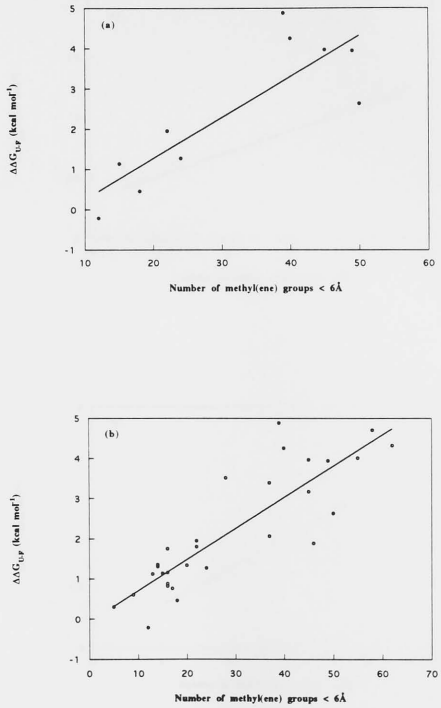


Fig. 5.4: Correlation between the number of side chain methyl(ene) groups within a 6Å radius of the group deleted from wild-type, and the change in free energy of unfolding for the mutations of hydrophobic residues in a) CI2 (slope = 0.10; R = 0.85) and b) CI2 and barnase (slope = 0.08; R = 0.87); filled circles (•) are CI2, open circles (o) are barnase.

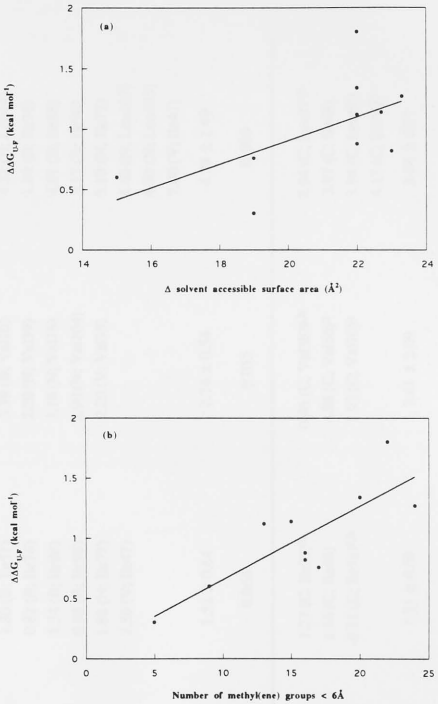


Fig. 5.5: Correlation between the change in free energy of unfolding for Ile to Val mutations in CI2 and barnase, and a) the difference in solvent accessible area that is buried by wild-type and mutant side chains (slope = 0.10; R = 0.60) and b) the number of side chains methyl(ene) groups within a 6Å radius of the group deleted from wild-type (slope = 0.06; R = 0.83).

Table 5.3: Calculation of the decrease in $\Delta\Delta G_{U\text{-}F}$ upon creating cavities in buried hydrophobic areas of barnase (B)[a], staphylococcal nuclease (N)[b], and bacteriophage f1 gene V protein (V)[c], compared with values for CI2 (C)[§]

	Ile ->Val (kcal mol^{-1})	Val ->Ala (kcal mol^{-1})	Ile ->Ala; Leu ->Ala (kcal mol^{-1})
	1.80 (B; Ile51)	3.39 (B; Val10)	4.32 (B; Leu14)
	0.82 (B; Ile76)	2.20 (N; Val39)	1.89 (B; Ile76)
	1.34 (B; Ile88)	3.10 (N; Val76)	4.01 (B; Ile88)
	0.88 (B; Ile96)	2.90 (N; Val104)	3.17 (B; Ile96)
	1.80 (N; Ile72)	2.20 (V; Val35)	5.10 (N, Ile72)
	2.50 (V; Ile47)		4.60 (N; Leu103)
			4.90 (N; Leu125)
			7.00 (V; Ile47)
Ave. $\Delta\Delta G_{U\text{-}F}$ (kcal mol^{-1})	1.52 ± 0.64	2.76 ± 0.54	4.38 ± 1.49
Ave. $\Delta\Delta G_{U\text{-}F}$ (kcal mol^{-1} Å^{-2}) [d]	0.066	0.055	0.060
	1.27 (C; Ile39)[h]	0.46 (C; Val38)[g,h]	2.64 (C; Leu27)[h]
	1.14 (C; Ile48)	4.88 (C; Val66)[h]	3.97 (C; Ile48)
	-0.21 (C; Ile76)[f,h]	1.93 (C; Val70)[h]	3.94 (C; Leu68)
			4.17 (C; Ile76)[h]
Ave. $\Delta\Delta G_{U\text{-}F}$(kcal mol^{-1})	1.21 ± 0.09	3.41 ± 2.09	3.68 ± 0.71
Ave. $\Delta\Delta G_{U\text{-}F}$(kcal mol^{-1} Å^{-2}) [e]	0.053	0.068	0.050

[a] Serrano *et al.*, 1992; [b] Shortle *et al.*, 1990; [c] Sandberg & Terwilliger, 1991; [d] solvent accessible areas of barnase & staph. nuclease, calculated using the algorithm of Lee & Richards (1971), are taken from Serrano *et al.*, 1992 (Table 5); [e] calculated as in d; [f] not incl. in ave. as more stable than wild-type CI2; [g] not included in average as Val38 not completely buried; [h] Jackson *et al.*, 1993a.

[§] adapted from Serrano *et al.*, 1992

Table 5.4: Comparison of the contribution of hydrophobic residues to protein stability with ΔG_{tr} values derived from model compounds

Mutant	% burial of side chain[a]	$\Delta\Delta G_{U\text{-}F}$<$m_{U\text{-}F}$> measured kcal mol^{-1}	$\Delta\Delta G_{U\text{-}F}$ [c] corrected for 100% burial kcal mol^{-1}	ΔG_{tr} [d] H_2O to *n*-octanol kcal mol^{-1}
IV48	92	1.14[b]	1.24	1.36
IA48	92	3.97[b]	4.32	3.84
LA68	100	3.94[b]	3.94	3.87
LA27	99	2.64[a]	2.67	3.87
VA38	56	0.46[a]	0.83	2.48
IV39	100	1.27[a]	1.27	1.36
VA66	100	4.88[a]	4.88	2.48
VA70	91	1.95[a]	2.14	2.48
IA76	100	4.25[a]	4.25	3.84

[a] Jackson *et al.* (1993a); [b] See Table 5.1; [c] $\Delta\Delta G_{U\text{-}F}$ values corrected so that all values are as if the mutated side chain were 100% buried; [d] ΔG_{tr} values based on measurements of transfer of amino acid side chains from water to *n*-octanol by Fauchere & Pilska (1983), corrected for the difference in volume between the solutes and solvents by Sharp *et al.* (1991).

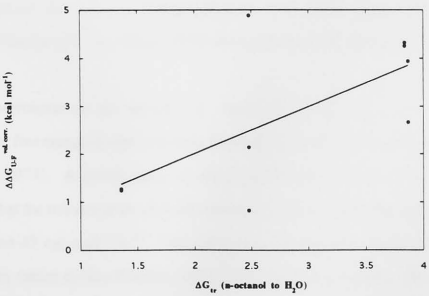


Fig. 5.6: Correlation between the change in free energy of unfolding for the mutations of CI2 hydrophobic core residues and the free energy of transfer of amino acids from *n*-octanol to water (slope = 0.99; r=0.67).

5.6 Discussion of Hydrophobic Core Results

The effect of cavity creating mutations on the stability of CI2 has been measured with great accuracy using both GdnHCl-induced equilibrium denaturation and calorimetry. From such studies, it is possible to estimate the effects on stability of deleting up to four methyl(ene) groups from within the core of the protein. Mutating Ile to Val (equivalent to deleting one methylene group) at position 48 costs the protein 1.14 kcal mol^{-1} in stabilisation energy, while further truncation to Ala (equivalent to deleting three methyl(ene) groups in total) destabilises CI2 by 3.97 kcal mol^{-1}. Truncation of Leu68 to Ala destabilises the protein by 3.94 kcal mol^{-1}, where such a mutation is equivalent to the removal of three methyl(ene) groups. Based on the above data in combination with that from analysis of other CI2 hydrophobic core mutants (see Table 5.2), the average change in free energy of unfolding (± standard deviation) for mutating an Ile to a Val is 1.18 ± 0.09 kcal mol^{-1}, for mutating Val to Ala is 3.42 ± 1.47 kcal mol^{-1}, while for mutating either an Ile or a Leu to an Ala is 3.64 ± 0.60 kcal mol^{-1}. On average, the deletion of one methyl(ene) group from the hydrophobic core of CI2 costs the protein 1.33 ± 0.45 kcal mol^{-1} in free energy of unfolding. Within the distribution of results, these values agree with those obtained from other studies of buried residues (summarised in Table 5.3), e.g., barnase (Serrano *et al.*, 1992), bacteriophage f1 gene V protein (Sandberg & Terwilliger, 1991) and staphylococcal nuclease (Shortle *et al.*, 1990).

Experiments on the transfer of nonpolar solutes have shown that the hydrophobic free energy is approximately proportional to the surface area of the solute (Hermann, 1977). A recent study on model compounds (Sharp *et al.*, 1991) has suggested that the transfer of an aliphatic hydrocarbon from a nonpolar solvent to water costs around 47 cal mol^{-1} $Å^{-2}$. This value is larger than the 20-30 cal mol^{-1} $Å^{-2}$ suggested by earlier studies (Chothia, 1976; Eisenberg & McLachlan, 1986; Ooi *et al.*, 1987), and includes a correction for the difference in volume between the solutes and solvents, previously neglected. A value of 47 cal mol^{-1} $Å^{-2}$ obtained by Sharp (1991) compares well with recent measurements on proteins (see Table 5.3). Analysis on

barnase (Kellis *et al.*, 1989), bacteriophage f1 gene V protein (Sandberg & Terwilliger, 1991) and staphylococcal nuclease (Shortle *et al.*, 1990) appear to show that the hydrophobic amino acids in the core of these proteins contributes 60 cal mol^{-1} $Å^{-2}$, on average, to these proteins' stability. The value for proteins can be derived by division of $\Delta\Delta G_{U\text{-}F}$ with the changes in the solvent accessible surface area for the hydrophobic side chains of Leu, Ile, Val and Ala in solution (Miller *et al.*, 1987). Such a calculation results in an average of 57 cal mol^{-1} $Å^{-2}$ for CI2 (see Table 5.3).

Several correlations have been sought and found between the change in free energy of protein unfolding, $\Delta\Delta G_{U\text{-}F}$, and various parameters that are indicative of the environment surrounding the mutated protein. A strong correlation is found between $\Delta\Delta G_{U\text{-}F}$ and the difference in side chain solvent accessible surface area buried between wild-type and mutant protein. For CI2, the best correlation is between $\Delta\Delta G_{U\text{-}F}$ and the number of methyl(ene) side chain groups that surround the deleted methyl(ene) groups in the mutant protein within a 6 Å radius. Similar correlations have been observed in barnase, suggesting that such environmental/free energy relationships may be general for proteins.

5.7 Structural Information on the Hydrophobic Core of Mutant Proteins

To interpret changes in stability upon mutation of hydrophobic core residues it is important to know the structural effects of the mutations. Structural and thermodynamic analysis of cavity-creating mutations has been performed using barnase (Buckle *et al.*, 1993) and T_4 lysozyme (Eriksson *et al.*, 1992b). In barnase, although the magnitude and nature of structural shifts depends on the site of mutations, some common features have emerged. The mutated side chain is found to move to the greatest extent, generally in the direction of the cavity created by mutation and no solvent molecules are found in any of the created cavities (Buckle *et al.*, 1993). No solvent molecules are found in the cavities created by mutating the hydrophobic core of T_4 lysozyme, either (Eriksson *et al.*, 1992b), as expected. Introduction of water into the hydrophobic interior of a protein would be energetically very unfavourable.

For T_4 lysozyme, there is a correlation between the change in free energy and size of cavity created upon mutation (Eriksson *et al.*, 1992b). The most destabilised hydrophobic core mutants are also those that have the largest cavities. In the case of marginally unstable mutants, repacking in the vicinity of the deletion is found to offset the energetic consequences of creating a cavity. No similar correlation is found for barnase mutants, although any correlation between cavity size and stability may be obscured as the cavity size and change in free energy of the barnase core mutants lie within a relatively narrow range (Buckle *et al.*, 1993).

In barnase, creating a cavity at the centre of a densely packed hydrophobic core (mutation IV88) results in side chain disorder. Free rotation of the Val88 side chain would only be possible if the surrounding structure were to "breathe" to some extent (Buckle *et al.*, 1993). Eriksson *et al.* (1992a) have shown that it is possible to place a benzene molecule inside an isolated cavity within the hydrophobic core of T_4 lysozyme. This illustrates the flexible nature of proteins, even in the hydrophobic core.

Structural analysis of the marginally stabilising CI2 core mutant, IV76 (on the edge of the core), reveals that there are small movements of side chains in the core, the largest of which (0.7 Å) is movement of the mutated side chain itself. It would appear that these small movements compensate, in part, for the cavity created by the mutation (Jackson *et al.*, 1993a). As is the case in barnase (Buckle *et al.*, 1993) and T_4 lysozyme (Erikkson *et al.*, 1992b), no solvent molecule occupies the cavity created by the mutation. Crystal structures have not yet been elucidated for other hydrophobic core mutants of CI2. Based on the structure of CI2 IV76, and hydrophobic core mutants of barnase (Buckle *et al.*, 1993) and T_4 lysozyme (Eriksson *et al.*, 1992b), however, it would appear that hydrophobic cores of proteins share some common features.

5.8 Conclusions

As first suggested by Kauzmann (1954, 1959) and others, hydrophobic interactions do play a very significant role in the stability of proteins. The result of

mutating hydrophobic core residues in CI2 shows this to be the case. The environment of the residue being mutated has a strong influence on the extent to which a particular mutation affects the protein's stability. Conventional measures of residue hydrophobicity cannot alone account for such changes. Correlations can be seen between the change in free energy of unfolding and parameters which are good indicators of residue environment. The changes in side chain solvent accessible area buried between wild-type and mutant and N (the number of methyl(ene) side chain groups that surround the deleted methyl(ene) groups in the mutant protein within a 6 Å radius) are both good indicators of how densely packed the environment of a mutated residue is. For mutations that occur in areas of high packing density, structural rearrangements are likely to be minimised due to the large number of contacts made amongst residues surrounding the site of mutation. For residues which lie towards the edge of the core, e.g., IV76 in CI2, structural rearrangements can occur on mutation which will result in a decrease in $\Delta\Delta G_{U\text{-}F}$. These results compare well with those of other protein studies (Shortle *et al.*, 1990; Serrano *et al.*, 1992a), implying that such correlations can be applied to the hydrophobic cores of globular proteins in general.

Chapter 6

Characterisation of the Transition State for CI2 Folding

6.1 Characterising the Transition State for Protein Folding

Fersht and coworkers (Matouschek *et al.*, 1989, 1990, 1992a, 1992b) have used a combination of protein engineering techniques and kinetics to analyse the interactions that are formed or broken in the intermediate or subsequent transition state for barnase folding. Individual mutations are positioned throughout the protein to act as probes for the formation of structure in the vicinity of the mutation. This method, in combination with NMR, has provided extensive structural characterisation of the folding intermediate of barnase, and the sequence of events in the folding pathway (Matouschek *et al.*, 1989, 1990). The theory, assumptions and limitations of this approach have been discussed in detail by Fersht and colleagues (Matouschek and Fersht, 1991; Fersht *et al.*, 1992), an outline of which is presented here.

The strategy used for the analysis of barnase may be applied to other proteins, and is as follows: (i) localised interactions in the native structure of the protein are disrupted by making conservative mutations, i.e., mutations that do not produce major structural changes; (ii) the effect of mutation on the free energy of unfolding is then measured by comparison with wild-type; (iii) the rates of unfolding and refolding of wild-type and mutant proteins are measured in order to determine the fraction of the change in free energy of unfolding used in stabilising the transition state (Matouschek *et al.*, 1989; Fersht *et al.*, 1992). A value ϕ (the ratio $\Delta\Delta G_{\ddagger}/\Delta\Delta G_{\text{U-F}}$) is used as a measure of the similarity between the structure of the transition state and that of the unfolded or native state of the protein. In the case where several non-covalent

interactions are made/broken, then ϕ is a measure of the weighted average change for all interactions. The theory applies best when fewer interactions are affected (Fersht *et al.*, 1992).

The major assumptions of the method are as follows: (i) the unfolding and refolding pathways are the same, (ii) mutation does not alter the pathway of folding, (iii) the structure of the folded state is not significantly changed by the mutation, (iv) the unfolded state is not perturbed by the mutation, and (v) the groups targeted for mutation do not make new interactions with new partners during the course of the reaction.

Unlike barnase, CI2 has been shown to follow a two-state model of protein folding both thermodynamically and kinetically, in which only one kinetically significant transition state exists (Jackson & Fersht, 1991a). This greatly simplifies analysis of kinetics data, since unfolding and refolding rates reflect simply the activation energies for the forward and reverse reactions. This provides a double check on measurements of $\Delta\Delta G_{\ddagger}$ and the individual ϕ values, which are ideally related by $\phi_u = 1 - \phi_f$. In this study, using the kinetics of unfolding and refolding of hydrophobic core and helix mutants, in conjunction with equilibrium studies on these mutants (discussed in Chapter 3), the nature of the transition state of CI2 is analysed.

6.2 Analysis and Interpretation of ϕ Values

Interpretation of ϕ values is simple in the case where $\phi = 1$ or $\phi = 0$. When $\Delta\Delta G_{\ddagger}$ is equal to zero or is much smaller than $\Delta\Delta G_{U\text{-}F}$ (see Fig. 6.1a) then $\phi = 0$. This suggests that interactions formed at the site of the mutation in the folded state of the protein are broken in the transition state. Likewise, when $\Delta\Delta G_{\ddagger}$ and $\Delta\Delta G_{U\text{-}F}$ are equal (see Fig. 6.1c) then $\phi = 1$, suggesting that interactions formed in the folded state of the protein at the site of the mutation are maintained in the transition state. When $\Delta\Delta G_{\ddagger}$ is smaller than $\Delta\Delta G_{U\text{-}F}$ (see Fig. 6.1b) then fractional values of ϕ result. Interpretation of fractional ϕ values is more difficult than interpretation when ϕ is either zero or one. Some general interpretations may be possible, however. In the first instance, fractional values of ϕ could indicate that there is partial formation of

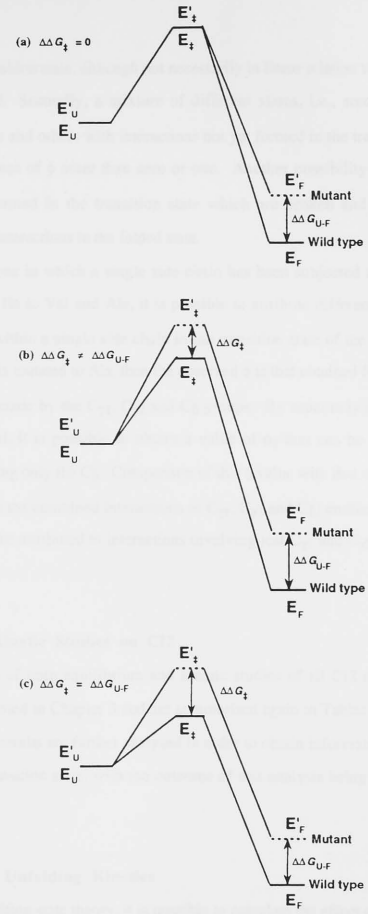


Fig. 6.1a - c: Free Energy Profile for the Unfolding of a Protein *via* 1 Major Transition State, where a) $\Delta\Delta G_{\ddagger} = 0$, b) $\Delta\Delta G_{\ddagger} \neq \Delta\Delta G_{U\text{-}F}$ and c) $\Delta\Delta G_{\ddagger} = \Delta\Delta G_{U\text{-}F}$

structure in the transition state, although not necessarily in linear relation to the value of ϕ that is measured. Secondly, a mixture of different states, i.e., some with fully formed interactions and others with interactions not yet formed in the transition state, could result in values of ϕ other than zero or one. Another possibility is that some interactions are formed in the transition state which are broken and replaced by different, stronger interactions in the folded state.

For mutations in which a single side chain has been subjected to a series of substitutions, e.g., Ile to Val and Ala, it is possible to attribute different ϕ values to particular groups within a single side chain in the transition state of the protein. For example, when Ile is mutated to Ala, then the observed ϕ is that obtained for an average of all interactions made by the $C_{\gamma 1}$, $C_{\gamma 2}$ and C_{δ} groups. By separately analysing the mutation Ile to Val, it is possible to obtain a value of ϕ_f that can be attributed to interactions involving only the C_{δ}. Comparison of this ϕ value with that of mutation to Ala (resulting from the combined interactions of $C_{\gamma 1}$, $C_{\gamma 2}$ and C_{δ}) enables calculation of a value for ϕ to be attributed to interactions involving just $C_{\gamma 1}$ and $C_{\gamma 2}$ (as in Table 6.5).

6.3 Results of Kinetic Studies on CI2

The results of both equilibrium and kinetic studies of all CI2 mutants have already been presented in Chapter 3 and are summarised again in Tables 6.1 - 6.3. In this chapter, these results are further analysed in order to obtain information about the structure of the transition state, with the outcome of this analysis being presented in Tables 6.4 and 6.5.

6.4 Analysis of Unfolding Kinetics

Using transition state theory, it is possible to calculate the effect of mutating a single or multiple residues in CI2 on the free energy of the transition state for unfolding. The stability of the transition state of mutant CI2, relative to wild-type, can

be calculated using equation 6.1, this being used to obtain information about the structure of the transition state:

$$\Delta\Delta G_{\ddagger\text{-F}} = -RT\ln\frac{k_u}{k_{u'}} \qquad (6.1)$$

where $\Delta\Delta G_{\ddagger\text{-F}}$ is the difference in free energy of the transition state for unfolding relative to the folded state between wild-type and mutant protein, and k_u and $k_{u'}$ are the rate constants of unfolding for wild-type and mutant, respectively. Values of $\Delta\Delta G_{\ddagger\text{-F}}$, calculated at both 0 M and 4 M GdnHCl, given in Table 6.2, are the same within error for most mutants (and within 10-20% for those outside of the error limits). It can be seen that the standard errors associated with values of $\Delta\Delta G_{\ddagger\text{-F}}$ calculated at 4 M ($\Delta\Delta G_{\ddagger\text{-F}}^{4M} \pm 0.06$ kcal mol^{-1}, on average) are much smaller than those for energy values calculated from the rate constants extrapolated back to water ($\Delta\Delta G_{\ddagger\text{-F}}^{H_2O} \pm 0.18$ kcal mol^{-1}, on average). The ratio of $\Delta\Delta G_{\ddagger\text{-F}}$ and $\Delta\Delta G_{\text{U-F}}$ is termed ϕ_u (see equation 6.2), and can be used as a measure of how fully interactions formed in the folded state of a protein are maintained in the transition state:

$$\phi_u = \frac{\Delta\Delta G_{\ddagger\text{-F}}}{\Delta\Delta G_{\text{U-F}}} \qquad (6.2)$$

where $\Delta\Delta G_{\text{U-F}}$ is the difference in the free energy of unfolding between wild-type and mutant ground states, as determined from GdnHCl-induced equilibrium denaturation experiments. ϕ_u, listed in Table 6.4 as $(1 - \phi_u)$, is calculated in two ways. In the first instance, $\phi_u^{H_2O}$ is calculated from a ratio of $\Delta\Delta G_{\ddagger\text{-F}}^{H_2O}$ to $\Delta\Delta G_{\text{U-F}}^{<m_{\text{U-F}}>}$, which results in values of $\phi_u^{H_2O}$ that have reasonable standard errors, for most mutants. The merits of using $\Delta\Delta G_{\text{U-F}}^{<m_{\text{U-F}}>}$ instead of $\Delta\Delta G_{\text{U-F}}^{H_2O}$ have already been discussed in Chapter 3, and is a reasonable approximation of $\Delta\Delta G_{\text{U-F}}$ for the purpose of calculating free energy differences in water. Secondly, $\phi_u^{4M\ Gdn}$ (listed in Table 6.4) is calculated from the ratio of $\Delta\Delta G_{\ddagger\text{-F}}^{4M}$ and $\Delta\Delta G_{\text{U-F}}^{4M}$. No approximations are used in calculating $\phi_u^{4M\ Gdn}$ as both ground and transition state energies are measured under

conditions where the unfolding reaction is being monitored, i.e., 4 M GdnHCl. Hence, $\phi_u^{4M\ Gdn}$ has direct bearing on the interactions maintained/broken in the transition state for unfolding. For the purpose of comparing ϕ_u with ϕ_f, however, values calculated in water are used. With few exceptions, all values of ϕ_u have values intermediate between zero and one (Table 6.4).

Table 6.1: Summary of GdnHCl-Induced Equilibrium data for Wild-type and Mutants of CI2[a]

Mutant	$\Delta\Delta G_{U\text{-}F}^{<m_{U\text{-}F}>}$ kcal mol^{-1}	$\Delta\Delta G_{U\text{-}F}^{4M\ Gdn}$ kcal mol^{-1}	$m_{U\text{-}F}$ kcal L mol^{-2}
wild-type			1.90 ± 0.03
SG31	0.82 ± 0.04	0.89 ± 0.09	2.10 ± 0.07
SA31	0.92 ± 0.11	0.86 ± 0.13	1.86 ± 0.08
EA33EA34	0.83 ± 0.10	0.82 ± 0.17	2.11 ± 0.18
SG31EA33EA34	1.67 ± 0.13	1.81 ± 0.15	2.15 ± 0.16
SA31EA33EA34	1.71 ± 0.13	1.88 ± 0.14	2.18 ± 0.14
EQ33	0.31 ± 0.05	0.30 ± 0.14	1.87 ± 0.10
ED33	0.54 ± 0.06	0.49 ± 0.11	1.83 ± 0.11
EN33	0.72 ± 0.06	0.73 ± 0.13	2.01 ± 0.10
EQ34	0.48 ± 0.05	0.45 ± 0.13	1.84 ± 0.09
ED34	0.76 ± 0.09	0.84 ± 0.17	2.20 ± 0.16
EN34	1.10 ± 0.07	1.21 ± 0.11	2.18 ± 0.11
IV48	1.14 ± 0.09	1.14 ± 0.09	1.99 ± 0.13
IA48	3.97 ± 0.08	4.14 ± 0.39	2.07 ± 0.19
LA68	3.94 ± 0.07	3.94 ± 0.25	1.98 ± 0.12
IA48IV76	4.17 ± 0.09	3.85 ± 0.22	1.83 ± 0.10

[a] See Table 3.1 for details.

Table 6.2: Summary of Unfolding Kinetics for Wild-type and Mutants of CI2 and Solvent Accessibility of the Transition State for Unfolding

Mutant	$m_{\ddagger\text{-}F}$ [a] M^{-1}	$\ln k_u^{H_2O}$ [a]	$\ln k_u^{4M\ Gdn}$ [a]	$\Delta\Delta G_{\ddagger\text{-}F}^{H_2O}$ [b] kcal mol^{-1}	$\Delta\Delta G_{\ddagger\text{-}F}^{4M\ Gdn}$ [c] kcal mol^{-1}	$\frac{m_U^{\ddagger}}{m_{U\text{-}F}}$ [d]
wild-type	1.31 ± 0.01	-9.04 ± 0.07	-3.90 ± 0.02	0.00	0.00	0.41 ± 0.02
SG31	1.30 ± 0.04	-8.42 ± 0.22	-3.24 ± 0.08	0.36 ± 0.23	0.39 ± 0.08	0.37 ± 0.05
SA31	1.29 ± 0.04	-8.53 ± 0.22	-3.37 ± 0.08	0.30 ± 0.23	0.32 ± 0.08	0.41 ± 0.06
EA33EA34	1.31 ± 0.04	-8.17 ± 0.23	-2.95 ± 0.08	0.52 ± 0.24	0.57 ± 0.08	0.37 ± 0.11
SG31EA33EA34	1.18 ± 0.05	-6.66 ± 0.28	-1.95 ± 0.09	1.41 ± 0.29	1.15 ± 0.09	0.33 ± 0.09
SA31EA33EA34	1.21 ± 0.02	-7.03 ± 0.10	-2.18 ± 0.03	1.19 ± 0.12	1.02 ± 0.04	0.33 ± 0.08
EQ33	1.32 ± 0.02	-8.97 ± 0.12	-3.70 ± 0.04	0.04 ± 0.14	0.12 ± 0.04	0.42 ± 0.07
ED33	1.33 ± 0.05	-8.76 ± 0.29	-3.43 ± 0.09	0.16 ± 0.30	0.28 ± 0.09	0.43 ± 0.09
EN33	1.38 ± 0.02	-9.01 ± 0.13	-3.51 ± 0.05	0.02 ± 0.15	0.24 ± 0.05	0.41 ± 0.07
EQ34	1.30 ± 0.03	-8.71 ± 0.16	-3.53 ± 0.05	0.20 ± 0.17	0.22 ± 0.06	0.42 ± 0.07
ED34	1.37 ± 0.01	-8.28 ± 0.08	-2.82 ± 0.03	0.45 ± 0.11	0.64 ± 0.03	0.37 ± 0.10
EN34	1.26 ± 0.03	-7.95 ± 0.10	-2.89 ± 0.06	0.65 ± 0.12	0.60 ± 0.07	0.34 ± 0.07
IV48	1.32 ± 0.01	-8.00 ± 0.09	-2.73 ± 0.03	0.62 ± 0.11	0.69 ± 0.04	0.39 ± 0.09
IA48	1.10 ± 0.04	-4.30 ± 0.18	0.10 ± 0.03	2.80 ± 0.19	2.37 ± 0.04	0.31 ± 0.10
LA68	1.15 ± 0.03	-6.03 ± 0.15	-1.46 ± 0.05	1.77 ± 0.17	1.44 ± 0.05	0.34 ± 0.07
IA48IV76	1.09 ± 0.03	-4.33 ± 0.15	0.04 ± 0.04	2.79 ± 0.17	2.34 ± 0.04	0.36 ± 0.06

[a] See Table 3.4 for details; [b] $\Delta\Delta G_{\ddagger\text{-}F}^{H_2O}$ calculated as follows: $\Delta\Delta G_{\ddagger\text{-}F}^{H_2O} = -RT\,(\ln k_u^{H_2O}{}_{wild\text{-}type} - \ln k_u^{H_2O}{}_{mutant})$; c $\Delta\Delta G_{\ddagger\text{-}F}^{4M\ GdnHCl}$ calculated as follows: $\Delta\Delta G_{\ddagger\text{-}F}^{4M} = -RT(\ln k_u^{4M}{}_{wild\text{-}type} - \ln k_u^{4M}{}_{mutant})$; [d] $m_U^{\ddagger}/m_{U\text{-}F} = RT(m_{\ddagger\text{-}F})/m_{U\text{-}F}$; where $m_{\ddagger\text{-}F}$ is given here in column 1 and $m_{U\text{-}F}$ in Table 6.1.

6.5 Analysis of Refolding Kinetics

The stability of the transition state of mutant CI2 relative to that of wild-type can be calculated using equation 6.3, this being used to obtain information about the structure of the transition state:

$$\Delta\Delta G_{\ddagger\text{-U}} = -RT\ln\frac{k_f}{k_f'} \qquad (6.3)$$

where $\Delta\Delta G_{\ddagger\text{-U}}$ is the difference in free energy of the transition state for folding relative to the unfolded state between wild-type and mutant CI2, and k_f and k_f' are the rate constants of folding for wild-type and mutant protein, respectively. $\Delta\Delta G_{\ddagger\text{-U}}^{H_2O}$ values, calculated from the combined GdnHCl-jump and pH-jump refolding kinetics (see Chapter 3), are presented in Table 6.3.

Similar to the analysis of the transition state for unfolding, the ratio ϕ_f (equation 6.4) is employed in order to analyse the structure of the transition state for folding:

$$\phi_f = \frac{\Delta\Delta G_{\ddagger\text{-U}}}{\Delta\Delta G_{\text{U-F}}} \qquad (6.4)$$

where $\Delta\Delta G_{\text{U-F}}$ is the difference in free energy of unfolding between mutant and wild-type CI2, as measured by GdnHCl-induced equilibrium denaturation. $\phi_f^{H_2O}$ values are listed in Table 6.4. Figures 6.2a - c show plots of ϕ_f values for different mutants versus the state of the protein, i.e., as the protein proceeds from the unfolded to the folded state via the transition state.

For the two-state model of protein folding, where only native and denatured states are significantly populated, just one kinetically important transition state exists (see Figs. 6.1a - c). Under these circumstances, ϕ may be discussed in terms of either unfolding or refolding. Thus, the strength of interactions in the transition state may be measured equally well via unfolding kinetics or via refolding kinetics to give values of ϕ_u and ϕ_f, respectively. In Chapter 3 it has been shown that mutants of CI2, like wild-

type (Jackson & Fersht, 1991a), follow a two-state model for protein folding, both thermodynamically and kinetically. This is a good indication that none of the mutations have a significant effect on the folding pathway of CI2.

Table 6.3: Summary of Refolding Kinetics for Wild-type and Mutants of CI2 and Solvent Accessibility of the Transition State for Refolding

Mutant	$m_{\ddagger\text{-U}}$ [a] M^{-1}	$\ln k_f^{H_2O}$ [a]	$\Delta\Delta G_{\ddagger\text{-U}}^{H_2O}$ [b] kcal mol^{-1}	$\frac{m_F^{\ddagger}}{m_{U\text{-}F}}$ [a,c]
wild-type	-1.90 ± 0.02	4.02 ± 0.02		0.59 ± 0.02
SG31	-2.01 ± 0.07	3.46 ± 0.07	0.33 ± 0.08	0.57 ± 0.05
SA31	-1.98 ± 0.05	3.25 ± 0.05	0.46 ± 0.06	0.63 ± 0.05
EA33EA34	-1.72 ± 0.02	3.06 ± 0.03	0.57 ± 0.03	0.49 ± 0.07
SG31EA33EA34	-1.77 ± 0.05	2.67 ± 0.04	0.80 ± 0.05	0.49 ± 0.07
SA31EA33EA34	-1.94 ± 0.04	2.78 ± 0.04	0.74 ± 0.04	0.53 ± 0.06
EQ33	-1.85 ± 0.03	3.30 ± 0.04	0.43 ± 0.05	0.59 ± 0.06
ED33	-1.95 ± 0.05	3.54 ± 0.07	0.28 ± 0.07	0.63 ± 0.07
EN33	-1.84 ± 0.04	3.00 ± 0.04	0.61 ± 0.04	0.55 ± 0.05
EQ34	-1.86 ± 0.04	3.36 ± 0.05	0.39 ± 0.05	0.60 ± 0.05
ED34	-1.90 ± 0.05	3.53 ± 0.06	0.29 ± 0.06	0.51 ± 0.07
EN34	-1.88 ± 0.04	2.93 ± 0.04	0.65 ± 0.05	0.51 ± 0.05
IV48	-2.00 ± 0.07	3.56 ± 0.06	0.27 ± 0.06	0.60 ± 0.07
IA48	-2.28 ± 0.09	2.25 ± 0.06	1.05 ± 0.06	0.65 ± 0.11
LA68	-2.2 ± 0.09	0.46 ± 0.06	2.11 ± 0.06	0.66 ± 0.08
IA48IV76	-2.66 ± 0.10	1.98 ± 0.05	1.21 ± 0.05	0.86 ±0.10

[a] from a linear fit of combined GdnHCl-jump and pH-jump refolding kinetic data, see Table 3.5; [b] calculated using combined GdnHCl-jump and pH-jump refolding kinetic data, as follows: $\Delta\Delta G_{\ddagger\text{-U}}^{H_2O} = -RT(\ln k_u^{H_2O}{}_{\text{wild-type}} - \ln k_u^{H_2O}{}_{\text{mutant}})$; [c] $m_F^{\ddagger}/m_{U\text{-}F} = -RT(m_{\ddagger\text{-U}})/m_{U\text{-}F}$, where $m_{\ddagger\text{-U}}$ is given here and $m_{U\text{-}F}$ can be found in Table 6.1.

In a two-state transition, it is possible to relate $\Delta\Delta G_{\ddagger\text{-}U}$ to $\Delta\Delta G_{\ddagger\text{-}F}$ and $\Delta\Delta G_{U\text{-}F}$ as shown in equation 6.5. For the purpose of this analysis ϕ is taken to be $\phi_f{}^{H_2O}$, either measured directly from refolding kinetics or in terms of $\phi_u{}^{H_2O}$, where ϕ_u is related to $\phi_f{}^{H_2O}$ by equation 6.6:

$$\Delta\Delta G_{\ddagger\text{-}U} = \Delta\Delta G_{U\text{-}F} - \Delta\Delta G_{\ddagger\text{-}F} \qquad (6.5)$$

$$\phi_u{}^{H_2O} = 1 - \phi_f{}^{H_2O} \qquad (6.6)$$

Table 6.4 : ϕ Values for Unfolding and Refolding of CI2

Mutant	[a](1-$\phi_u{}^{H_2O}$)	[b](1-$\phi_u{}^{4M\ Gdn}$)	[c]$\phi_f{}^{H_2O}$
SG31	0.56 ± 0.28	0.56 ± 0.13	0.41 ± 0.09
SA31	0.67 ± 0.25	0.63 ± 0.18	0.50 ± 0.09
[d]SG31EA33EA34	-0.06 ± 0.02	0.45 ± 0.04	0.27 ± 0.05
[d]SA31EA33EA34	0.24± 0.14	0.54 ± 0.04	0.19 ± 0.04
[e]*(EQ33*	*0.87 ± 0.45*	*0.60 ± 0.34*	*1.37 ± 0.27)*
ED33	0.70 ± 0.56	0.43 ± 0.17	0.52 ± 0.15
EN33	0.97 ± 0.21	0.67 ± 0.18	0.84 ± 0.09
EQ34	0.58 ± 0.36	0.51 ± 0.20	0.82 ± 0.13
ED34	0.41 ± 0.16	0.24 ± 0.05	0.38 ± 0.09
EN34	0.41 ± 0.12	0.50 ± 0.07	0.59 ± 0.06
IV48	0.46 ± 0.11	0.39 ± 0.04	0.24 ± 0.06
IA48	0.29 ± 0.05	0.43 ± 0.04	0.26 ± 0.02
LA68	0.55 ± 0.04	0.63 ± 0.05	0.53 ± 0.02
[f]IA48IV76	1.00 ± 0.04	0.89 ± 0.02	0.80 ± 0.01

[a] $\phi_u{}^{H_2O} = \frac{\Delta\Delta G_{\ddagger\text{-}F}{}^{H_2O}}{\Delta\Delta G_{U\text{-}F}{}^{<m_{U\text{-}F}>}}$; [b] $\phi_u{}^{4M} = \frac{\Delta\Delta G_{\ddagger\text{-}F}{}^{4M}}{\Delta\Delta G_{U\text{-}F}{}^{4M\ Gdn}}$; [c] $\phi_f{}^{H_2O} = \frac{\Delta\Delta G_{\ddagger\text{-}U}}{\Delta\Delta G_{U\text{-}F}{}^{<m_{U\text{-}F}>}}$

[d] relative to double mutant EA33EA34 so that triple mutants SG31EA33EA34 and SA31EA33EA34 act as probes for SG31 and SA31, respectively; [e] i$\Delta\Delta G_{U\text{-}F}$ too small, resulting in ϕ values (italicised) difficult to interpret due to large errors; [f] relative to IA48 so that IA48IV76 acts as a probe for IV76.

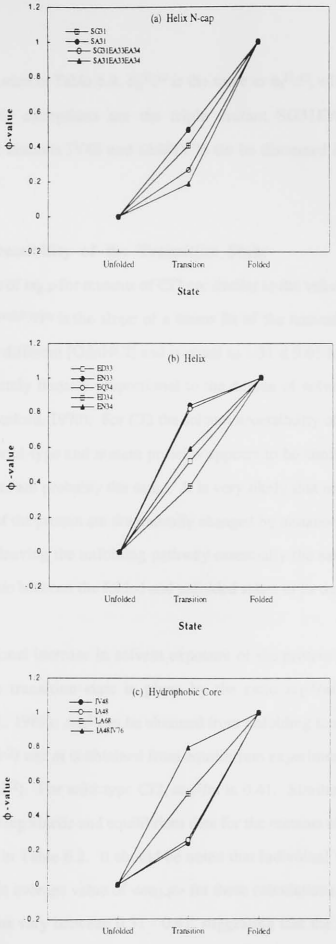


Fig. 6.2a - c: ϕ Plots for (a) the Helix N-cap, (b) the Helix N-cap+2 and N-cap+3 positions, and (c) the Hydrophobic Core.

As can be seen in Table 6.4, $\phi_u^{H_2O}$ is the same as $\phi_f^{H_2O}$, within error, for most mutants. Notable exceptions are the triple mutant SG31EA33EA34 and the hydrophobic core mutants IV48 and IA48IV76 (to be discussed in sections 6.7 and 6.8).

6.6 Solvent Accessibility of the Transition State

The values of $m_{\ddagger\text{-}F}$ for mutants of CI2 are similar to the value obtained for wild-type, where $m_{\ddagger\text{-}F}^{\text{wild-type}}$ is the slope of a linear fit of the natural logarithms of the unfolding rates in different [GdnHCl] and is equal to 1.31 ± 0.01 M^{-1} (see Table 6.2). $m_{\ddagger\text{-}F}$ is approximately linearly proportional to the degree of solvent exposure of the transition state (Tanford, 1970). For CI2 the solvent accessibility of the transition state for unfolding of wild-type and mutant proteins appears to be similar, suggesting that the transition states are probably the same. It is very likely that neither the transition nor folded states of the protein are dramatically changed by mutations made in the helix and core of CI2, leaving the unfolding pathway essentially the same. The transition state is intermediate between the folded and unfolded states in its degree of exposure to solvent.

The fractional increase in solvent exposure of the protein on going from the native to the transition state is given by the ratio $m_U^{\ddagger}/m$ (Tanford, 1970; Matouschek, *et al.*, 1989). $m_U^{\ddagger}$ can be obtained from unfolding kinetics ($m_{\ddagger\text{-}F}^{\text{w.t.}} \times RT$ = 0.77 kcal L mol^{-2}) and m is obtained from equilibrium experiments ($m_{U\text{-}F}^{\text{w.t.}}$ = 1.90 ± 0.03 kcal L mol^{-2}). For wild-type CI2, $m_U^{\ddagger}/m$ is 0.41. Similar calculations have been performed using kinetic and equilibrium data for the mutants of CI2, the results of which are shown in Table 6.2. It should be noted that individual values of $m_{U\text{-}F}$ are used rather than an average value of $\langle m_{U\text{-}F} \rangle$ for these calculations. Values of $m_U^{\ddagger}/m$ for mutant proteins vary between 0.31 - 0.41, suggesting that the transition state for unfolding of these mutants is similarly solvent exposed to the transition state of wild-type, relative to the unfolded state. In a similar manner, $m_F^{\ddagger}/m$, the fractional increase in exposure of the protein to solvent on going from the unfolded to the transition state,

may be calculated. For wild-type CI2, this value is 0.59 and is equal to (1 - $m_U^{\ddagger}/m$), as would be expected for a two-state transition. Values of $m_F^{\ddagger}/m$ determined for mutants of CI2 are listed in Table 6.3 are the same, within error, as (1 - $m_U^{\ddagger}/m$) for virtually all mutants.

6.7 ϕ Value Analysis of Mutations at the α-Helix N-terminus

For mutations that cause non-disruptive deletions with access of water at the site of mutation, e.g., mutations at the helix N-terminus, fractional values of $\phi_f^{H_2O}$ are difficult to interpret in a straightforward manner. This is because solvation of the mutated side chain can vary throughout the course of the reaction. Fractional values of $\phi_f^{H_2O}$ under these circumstances may be interpreted generally as a measure of how well interactions formed in the folded state of the protein are maintained or broken in the transition state. Alternatively, fractional $\phi_f^{H_2O}$ values may be the result of a mixture of different states.

(i) SA31 and SG31:

Ser31 is the N-cap residue of the α-helix in CI2, and participates in the formation of a hydrogen bond between its Oγ group and the backbone NH group of N-cap+3 residue Glu34. The side chain of Ser31 is on the solvent exposed face of the helix. Mutation of Ser31 to Gly (see Table 6.4) results in a $\phi_f^{H_2O}$ value of 0.41, suggesting that about 60 % of the interaction energy between residues in the helix is lost when going from the native to the transition state of the protein. Likewise, mutation of Ser31 to Ala results in a fractional $\phi_f^{H_2O}$ value of 0.50 (Table 6.4), similar to the value obtained for the mutant SG31. From such results, it is reasonable to conclude that the N-cap of the α-helix is partially formed in the transition state of CI2.

(ii) SG31EA33EA34 and SA31EA33EA34:

$\phi_f^{H_2O}$ calculated for the triple mutants SG31EA33EA34 and SA31EA33EA34 relative to the double mutant EA33EA34 will give information about the extent to which

N-cap interactions are formed in the transition state of a destabilised mutant of CI2. For the mutant SG31EA33EA34, the ϕ value varies between -0.06 - 0.45, depending on the method of calculation (see Table 6.4). $\phi_f^{H_2O}$ for SG31EA33EA34 does compare well, however, with $\phi_f^{H_2O}$ obtained for the triple mutant SA31EA33EA34 (0.27 and 0.19, respectively).

The values of $\phi_f^{H_2O}$ for the N-cap relative to the destabilising double mutant are somewhat lower than those obtained directly for both SA31 and SG31 single mutants. Similar observations have been made for destabilised helix mutants in the protein barnase (J. Matthews, unpublished data). Mutation of Ser to Gly in CI2 gives a $\phi_f^{H_2O}$ of 0.27 (*c.f.* 0.41 for the single mutant), while mutation of Ser to Ala reduces the $\phi_f^{H_2O}$ value to 0.19 (*c.f.* 0.50 for the single mutant). This would suggest that interactions at the helix N-cap are less fully formed in the transition state of the destabilised mutant EA33EA34 than they are in wild-type. Either fewer interactions are formed in the transition state of EA33EA34, or some interactions are weaker than their counterparts in wild-type. Another interpretation of these results is that the helix is fully formed slightly later along the folding pathway of the destabilised EA33EA34 mutant of CI2 than it would be in wild-type protein, which would account for the lower $\phi_f^{H_2O}$ values for the triple compared to the single N-cap mutants. Furthermore, evidence from comparisons of $m_{\ddagger\text{-}F}$ with $m_{U\text{-}F}$ for mutants of barnase suggest that movement of the reaction coordinate for the transition state can take place upon mutation (Matouschek & Fersht, 1993).

Because there is no structural evidence for direct contacts between Ser31 and the side chain of Glu33 or Glu34 in native wild-type CI2, these results may be an indication that part of the interaction energy for Ser31 in the wild-type protein is due to electrostatic interactions between the OH dipole of Ser31 and the charges of Glu33 and Glu34. The fact that such dipole-charge interactions cannot occur in the transition state of the double mutant (where both Glu33 and Glu34 are replaced with Ala) could account for the difference in SA31 and SG31 $\phi_f^{H_2O}$ values between the wild-type and EA33EA34. Although this contradicts the assumption that mutations do not affect the

pathway of folding, it is possible that the combined effect of the Glu to Ala mutations slows the process of N-cap formation or helix propagation at the N-cap. Given the importance of charge-dipole interactions to the stability of helices (Hol, 1978, 1985) it seems likely that the kinetics of helix formation is also affected by electrostatics. This may be manifested in different helix propagation propensities at positions with variously charged residues, as suggested by Kallenbach and coworkers (Gans *et al.*, 1991).

(iii) EQ33, ED33, EN33 and EQ34, ED34, EN34:

The side chains of Glu33 and Glu34, the third and fourth residues in the α-helix of CI2, are both fully solvated. Mutations of these side chains to Gln, Asp, and Asn provides multiple probes for helix formation. For the mutation Glu33 to Gln, ϕ_f values obtained from either unfolding or refolding kinetics must be interpreted cautiously as large errors are associated with each value. However, when comparing the $\phi_f^{H_2O}$ values for the mutations E33Q, E33D and E33N with E34Q, E34D, and E34N, it is possible to see that a similar trend occurs at both positions. For EQ33 a $\phi_f^{H_2O}$ value of roughly one is obtained. This suggests that interactions involving this residue are fully or nearly maintained in the transition state of the mutant. For ED33 a $\phi_f^{H_2O}$ value of 0.52 is obtained, suggesting that up to half of the interactions are lost on going from the folded to the transition state of this mutant. The mutation EN33 has a value of $\phi_f^{H_2O}$ which is somewhat lower than that observed for mutating Glu to Gln, but slightly higher than for mutating Glu to Asp at the position 33. A $\phi_f^{H_2O}$ value of 0.84 for EN33 is calculated, indicating that most of the interactions are maintained in the transition state. For EQ34, a $\phi_f^{H_2O}$ of 0.82 is observed compared with 0.38 for ED34 and 0.59 for EN34.

Interactions in the transition state of CI2 appear to be progressively weakened on mutation of Glu to Gln, Asn and then Asp, at both positions 33 and 34. It would seem that charge plays a less important role than the length of the side chain in determining to what extent interactions are formed in the transition state of the protein.

It may be that in a partially formed helix, charge-dipole interactions are not yet realised, and cannot be until the helix is fully folded and the necessary charge distribution is obtained to create a helix dipole.

6.8 ϕ Value Analysis of Mutations in the Hydrophobic Core

For mutations that cause non-disruptive deletions without access of water at the site of mutation, e.g., mutations in the hydrophobic core, fractional values of $\phi_f^{H_2O}$ may possibly be interpreted. For such hydrophobic to hydrophobic mutations $\phi_f^{H_2O}$ is a measure of the fractional loss in van der Waals interaction energies, with the assumption that the difference in solvation energy of the wild-type and mutant side chains are approximately the same (Matouschek *et al.*, 1989; Fersht *et al.*, 1992).

Values of ϕ obtained from different calculations are not the same, within error, for the hydrophobic core mutants IV48 and IA48IV76 (the latter being used as a probe for IA48 and IV76). For IA48IV76, $(1 - \phi_u^{H_2O})$ is the same as $\phi_f^{H_2O}$, within 20 %. For the mutant IV48, the difference between the two values is greater than this. The latter mutant does have intermediate values of both $\phi_u^{H_2O}$ and $\phi_f^{H_2O}$, so interpretation of this result is effectively the same whether looking at $\phi_u^{H_2O}$ or $\phi_f^{H_2O}$ (see discussion below).

(i) IV48 and IA48:

The side chain of Ile48, the second residue of β-strand 2, is completely buried in the centre of the core region of CI2. Mutating Ile to Val results in deletion of the C_δ methyl group of the Ile side chain and in so doing, removes interactions with the side chains of helix residues Lys36, Ile39 and Leu40. In this way, the mutation IV48 acts as a superb probe for interactions between the α-helix and second β-strand in CI2. A $\phi_f^{H_2O}$ value of 0.24 is obtained for the mutant IV48, suggesting that nearly 80 % of the interaction energy between the helix and β-strand 2 is lost in the region of this probe during unfolding of the native state to the transition state.

Further mutation of Ile48 to Ala deletes $C_{\gamma 1}$ and $C_{\gamma 2}$ as well as C_{δ} methyl(ene) groups, removing additional interactions with helix residues Lys36 and Ile39, as well as disrupting interactions with core residue Leu68 in β-strand 3 of the protein. Further to this, interactions between Ile48 and Ala46 (in the turn between α-helix and the 2nd β-strand) and Val50 (β-strand 2) are lost. A $\phi_f^{H_2O}$ value of 0.26 for the mutation IA48 is the same, within error, as $\phi_f^{H_2O}$ obtained for the mutant IV48, suggesting that no further interaction energy is lost in this region upon deletion of the $C_{\gamma 1}$ and $C_{\gamma 2}$ groups in Ile48. Thus, the mutation IA48 is dominated by a decrease in interaction energy between the α-helix and β-strand 2, while loss of other interactions goes unnoticed.

A fine structure φ analysis can be performed on the side chain of Ile48 in order to look at the mutation VA48 (see Table 6.5). Such a mutation can report on the effects of losing interactions specifically between the $C_{\gamma 1}$ and $C_{\gamma 2}$ groups of residue 48 and side chain atoms in the α-helix and β-strand 3 of CI2. A $\phi_f^{H_2O}$ value of 0.28 for the composite mutant VA48 is similar to those values observed for IV48 and IA48, suggesting that interactions in the transition state are formed to approximately the same extent by the different methyl(ene) groups of the Ile48 side chain.

(ii) LA68:

The central residue of the third β–strand is Leu68, the side chain of which is completely buried and lies in the centre of the protein's hydrophobic core. The $C_{\gamma 1}$ and $C_{\gamma 2}$ as well as $C_{\delta 1}$ methyl(ene) groups of Leu68 form interactions with α-helix residues Val32, Ala35 and Ile39, β-strand 2 residues Ile48 and Val50, and β-strand 4 residues Ile76 and Pro80. Mutation of Leu68 to Ala disrupts interactions between the β-sheet as well as between β-strand 3 and the α-helix. In this way, the mutation LA68 is a very good probe of the formation of the centre of the hydrophobic core as well as the interaction between the α-helix and β-strands 2, 3 and 4. A $\phi_f^{H_2O}$ value of 0.53 is obtained for the mutant LA68. This would suggest that around 50 % of the interaction energy between β-strand 3 and the β-sheet and α-helix is lost in the region of this probe during unfolding of the native state to the transition state, suggesting that the

β-sheet and centre of the helix are partly formed in the transition state of CI2. It may be possible that both the β-sheet and α-helix are formed to a large extent in the transition state, but interaction between the two appears to be weaker in the transition state than in the native state because the (largely formed) helix and sheet have not yet come together through close packing.

Table 6.5: Fine Structure ϕ Analysis of IV48 and IA48 Mutations

	Unfolding Kinetics [a]				
Mutant	$m_{‡\text{-}F}$ M^{-1}	$\ln k_u^{H_2O}$	$\Delta\Delta G_{‡\text{-}F}^{H_2O}$ kcal mol^{-1}	$\Delta\Delta G_{U\text{-}F}$ [b] kcal mol^{-1}	$\phi_u^{H_2O}$ [c]
IV48	1.32 ± 0.01	-8.00 ± 0.09	0.62 ± 0.11	1.14 ± 0.09	0.46 ± 0.09
IA48	1.10 ± 0.04	-4.30 ± 0.18	2.80 ± 0.19	3.97 ± 0.08	0.29 ± 0.02
VA48			2.18 ± 0.22	2.83 ± 0.12	0.23 ± 0.03
	Refolding Kinetics [d]				
	$m_{‡\text{-}U}$ M^{-1}	$\ln k_f^{H_2O}$	$\Delta\Delta G_{‡\text{-}U}^{H_2O}$ kcal mol^{-1}	$\Delta\Delta G_{U\text{-}F}$ [b] kcal mol^{-1}	$\phi_f^{H_2O}$ [c]
IV48	2.00 ± 0.07	3.56 ± 0.06	0.27 ± 0.06	1.14 ± 0.09	0.24 ± 0.06
IA48	2.28 ± 0.10	2.25 ± 0.06	1.05 ± 0.06	3.97 ± 0.08	0.26 ± 0.02
VA48			0.78 ± 0.08	2.83 ± 0.12	0.28 ± 0.03

[a] See Table 6.2; [b] $\Delta\Delta G_{U\text{-}F} = \Delta\Delta G_{U\text{-}F} < m_{U\text{-}F} >$, see Table 3.1 for details; [c] see Table 6.4; [d] see Table 6.3

(iii) IA48IV76, IA48 and IV76:

$\phi_f^{H_2O}$ calculated for the double mutant IA48IV76 relative to the single mutant IA48 (see Table 6.4) will give information about the extent to which interactions are formed in the transition state of the single mutant IV76. IV76 is located in the fourth strand of β-sheet, and lies on the edge of the hydrophobic core. The C_δ methyl group of Ile76 interacts with residues Leu68 and Val70 in β-strand 3, and Val32 and Ala35 in the α-helix. Thus, mutation of Ile76 to Val acts as a probe of the interactions formed by β-strand 4 with β-strand 3 and α-helix. A $\phi_f^{H_2O}$ of 0.80 for the double mutant (with respect to IA48) suggests that interactions formed in the native state by the side chain of Ile76 are almost fully maintained in the transition state. This value of $\phi_f^{H_2O}$ compares well with the value determined directly for the single mutant IV76 ($\phi_f^{H_2O}$ = 0.76 ± 0.60; Jackson *et al.*, 1993b). The error associated with $\phi_f^{H_2O}$ measured directly for IV76 is very high (due to the fact that both $\Delta\Delta G_{\ddagger\text{-}U}^{H_2O}$ and $\Delta\Delta G_{U\text{-}F}^{H_2O}$ for IV76 have large errors), making interpretation of this single result difficult. The combined results for IA48IV76 and IA48, however, do provide a similar $\phi_f^{H_2O}$ value to that obtained for IV76 alone, but with a much smaller error. Thus, an excellent probe for the interaction between the fourth β-strand, β-strand 3 and the α-helix is available.

By calculating $\phi_f^{H_2O}$ for the double mutant with respect to the single mutant IV76, it is possible to obtain information about the single mutant IA48. $\phi_f^{H_2O}$ calculated in this way is 0.24 (Jackson *et al.*, 1993b), which is comparable to the $\phi_f^{H_2O}$ value 0.26 obtained directly for the single IA48 mutant (Table 6.4). This suggests that the interaction between β-strand 2, β-strand 3 and the α-helix is weakened by at least 70% in the transition state compared to the native state in this region of the protein.

6.9 Discussion

(i) Validity of Extrapolating Unfolding Kinetic Data from GdnHCl to Water:

The kinetics of wild-type CI2, studied over the range 0 - 7 M GdnHCl, fit a two-state transition for protein folding, in which only one kinetically significant transition state exists, as stated previously (Jackson & Fersht, 1991a). The same holds

true for mutants of CI2, as discussed in Chapter 3 of this thesis. It is therefore possible to study the transition state directly from either direction of the folding reaction, i.e., from the folded state of the protein using unfolding kinetics, or from the unfolded state using refolding kinetics.

The principle of microscopic reversibility holds that a reaction and its reverse occur by the same pathway or mechanism if the two occur under identical conditions. Since the unfolding and refolding kinetics of CI2 are studied at high and low denaturant concentration, respectively, the assumption of reversibility cannot necessarily be made. Of great significance is the ability to measure changes in the energy of the transition state in the absence of denaturant, using data from the refolding kinetics in water. Comparison of the resulting $\phi_f^{H_2O}$ values with those calculated by extrapolating the unfolding kinetics to 0 M GdnHCl (given as $(1 - \phi_u^{H_2O})$) tests the assumption that refolding proceeds by the reverse route for unfolding, i.e., the refolding and unfolding pathways are the same. For the majority of mutants, the $\phi_f^{H_2O}$ values measured from refolding kinetics are the same, within experimental error, as those measured from unfolding kinetics. The validity of linearly extrapolating unfolding data measured in high concentrations of denaturant back to water has been discussed elsewhere (Fersht *et al.*, 1990).

(ii) Assumptions and Validity of ϕ Value Analysis:

Mutations in the hydrophobic core of CI2 and helix N-terminus were designed to cause minimal structural rearrangements in the folded protein. Small non-disruptive mutations, such as Ile to Val in the hydrophobic core and Ser to Ala and Gly at the helix N-terminus, destabilise CI2 by less that 2 kcal mol^{-1}, and do not introduce any radical new functional groups that are likely to make new interactions in the protein. Structural studies of more destabilising Ile/Leu to Ala mutants suggest that deletion of methyl(ene) groups creates cavities in the core rather than causing gross structural rearrangements and formation of new interactions. Mutations in the hydrophobic core and helix of CI2 provide multiple probes for each region of the protein. In the case of mutations IV48

and IA48, probes at the same position result in similar values of $\phi_f^{H_2O}$. Such consistency of ϕ values gives credence to the assumption that for most mutants, target groups do not make new interactions with new partners during the course of the reaction, and that the folding pathway has not been significantly changed upon mutation.

Equilibrium and kinetic studies suggest that none of the mutants deviate from the two-state model for folding. In addition, the solvent accessibility of the transition state is similar for wild-type and mutant CI2. This would suggest that both the transition and ground states of the protein are not dramatically altered by mutations, leaving the folding pathway essentially the same for the majority of mutants. Possible exceptions to this are the triple helix mutants, SG31EA33EA34 and SA31EA33EA34, which act as probes for the single mutants SG31 and SA31, respectively. In these cases, formation of the helix N-cap may possibly be delayed in the destabilised mutant EA33EA34 relative to the wild-type protein. Structural studies of these mutants indicate that there are no major structural rearrangements in the protein (Harpaz *et al.*, manuscript in preparation).

(iii) Structure of the Transition State:

Mutation of the hydrophobic core of CI2 provides information about the extent to which interactions that are formed between the α-helix of CI2 and various strands of β-sheet in the native state are maintained in the transition state. Mutation of the helix N-terminus provides direct information on the extent to which the helix itself is formed in the transition state of the protein. All mutations result in fractional values of ϕ (Table 6.4), suggesting that both the centre of the hydrophobic core and the N-terminus of the α-helix are partially formed in the transition state compared to the native state of the protein. It seems likely that the centre of the core has not yet attained the closely packed conformation that exists in the native structure. It is possible that the free energy of the hydrophobic core may be reduced as the number of van der Waals contacts in the core

decreases. In addition, the core may have become increasingly solvent exposed, which would also reduce the energy in the transition state compared with the native state.

Based on N-cap data alone, it is not possible to establish whether the helix is fully formed before or after consolidation of the hydrophobic core. It is necessary to obtain ϕ values for mutations throughout the helix, using as probes only those residues that form interactions within the helix itself, i.e., fully solvent exposed and preferably non-charged side chains.

6.10 Conclusions

There is strong evidence to indicate that the principle of microscopic reversibility applies to the system of CI2 folding, and that the unfolding and refolding pathways are the same. Mutation appears not to affect the folding pathway, nor significantly perturb the structure of either the transition or ground states of the majority of CI2 mutants. The transition state of CI2 is partially solvent exposed, wild-type and mutants experiencing a similar degree of solvent exposure. This is further indication that mutation does not significantly affect the pathway of folding in most cases. ϕ value analysis suggests that both the helix and hydrophobic core of CI2 are partially formed in the transition state. More work needs to be done in order to fully characterise the transition state for CI2, and determine the likely sequence of folding events which result in the native structure of the protein.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

The thermodynamics and folding kinetics of helix N-terminus and hydrophobic core mutants of CI2 have been studied, and it has been shown that all mutants adhere to a two-state model for protein folding and destabilise the protein relative to wild-type. Mutation of Ser at the N-cap of CI2 destabilises the protein by nearly 1 kcal mol^{-1}, although there is no apparent preference for Gly *versus* Ala at this position, with respect to both the wild-type and double mutant EA33EA34. Fulfilment of hydrogen-bonding requirements at the N-cap is important, however. Mutation of Glu33 at the N-cap+2 and Glu34 at the N-cap+3 to Gln Asp and Asn progressively destabilises the protein from 0.3 - 1.1 kcal mol^{-1}. For these residues, side chain length appears to affect the preference for certain amino acids more than does the presence of a negative charge. Statistical survey implies that, in general, electrostatic/dipole effects are very important in determining which side chains are favourable at the termini of helices. In CI2 the full effect of such electrostatic interactions may be partially screened due to the presence of GdnHCl in equilibrium denaturation studies, obscuring the true relative preferences for charged side chains at the helix N-terminus.

The effect of cavity creating mutations in the hydrophobic core of CI2 on stability has also been measured. Deletion of one methyl(ene) group from the core destabilises the protein by 1.3 kcal mol^{-1} on average, with a strong correlation between the environment of the mutation and its effect on stability. For mutations that occur in areas of high packing density, drastic structural rearrangements are

unlikely due to the large number of contacts made amongst residues surrounding the site of mutation. Thus, most of the destabilisation observed in core mutants has been attributed to a loss of favourable van der Waals energy in the folded state.

Finally, the unfolding and refolding kinetics of mutants have been studied, and it would seem that the unfolding and refolding pathways are the same. For most mutants, neither the folding pathway nor the structure of ground and transition states appears to be perturbed by mutation. ϕ value analysis suggests that both the helix N-terminus and hydrophobic core are partially formed in the transition state of CI2, being more exposed to solvent than the native state.

7.2 Future Work

(i) Circular Dichroism Denaturation Studies of Charge Mutants

To address the problem of potential screening effects of GdnHCl in equilibrium denaturation studies of CI2 mutants, thermal denaturation experiments in the absence of chemical denaturant (measured by calorimetry) were performed. As already stated in Chapter 3, due to undesired protonation of the charged side chains of Glu33 and Glu34 such experiments have proven fruitless in addressing the issue of how mutations that remove charge at these positions affect the stability of CI2. Attempts to perform calorimetry at a pH above the pK_a of Glu33 and Glu34, but close to the pI of the protein, failed due to aggregation. It may be possible to avoid such aggregation by using very dilute concentrations of protein.

The minimum protein concentration required to measure thermal denaturation by calorimetry is still very likely to be high enough to cause aggregation at or near the pI of CI2. It may be possible to monitor thermal denaturation of mutants, however, by circular dichroism (CD). The concentration of protein required for such experiments is about tenfold less than that required for calorimetry, and so CD experiments performed near the pI of CI2 may not be hampered by the problem of aggregation. CD-monitored thermal denaturation in the absence of GdnHCl, performed at a pH above the pK_a of Glu33 and Glu34 side chains, may provide a

means of measuring the true effect that mutation at these positions has on CI2 stability.

(ii) Full Characterisation of the Transition State of CI2

The mutants studied here provide information which has been used to partially characterise the transition state for CI2 folding. More work needs to done in order to fully characterise the transition state, and determine the likely sequence of folding events which result in the native structure of the protein. Based on the N-cap data alone, it is not possible to establish whether the helix is formed fully before of after consolidation of the hydrophobic core. It will be necessary to mutate additional residues that form interactions only within the helix itself, preferably at solvent exposed positions and with non-charged side chains. These mutants can be used as additional probes of helix-formation. At this time, site-directed mutants are being made throughout the helix and β-sheet of CI2 (L. Itzhaki and D. Otzen, unpublished data). The results of analysis on these mutations, in conjunction with the analysis of mutations presented in this study, will allow for a more complete characterisation of the transition state and folding pathway of CI2.

(iii) The Relationship between Structure and Function in Mutants of CI2

Structural studies on mutants of CI2 (Jackson *et al.*, 1993a; Harpaz *et al.*, manuscript in preparation) suggest that conservative mutations do not cause gross structural rearrangement within the protein, and that if rearrangements do occur, they tend to be small and localised in the vicinity of the mutated residue. Ideally, structures should be determined for all CI2 mutants in order to confirm that mutation does not greatly perturb the native state of the protein. If it is assumed that no global rearrangements have occurred in the protein as a result of mutation, then it should also be true that the structure of the reactive site loop, and hence, the inhibitory effects of CI2, on subtilisin for example, should not be affected by mutation.

A preliminary test of the effect CI2 mutants have on the protein's global structure (providing mutants do not directly involve the reactive site loop) would be to measure the ability of CI2 mutants to inhibit subtilisin. In the event that a particular mutation does affect the inhibitory function of CI2, then such a mutant would be a prime candidate for further structural studies by crystallography and/or NMR. In the case where global rearrangements have occurred, it would be of interest to determine why one mutation has the capacity to affect the protein's structure so dramatically while others leave the overall structure unchanged.

References

Alber, T., Dao-Pin, S., Wilson, K., Wozniak, J.A., Cook, S.P. & Matthews, B. (1987) *Nature* **330**, 41

Alber, A. (1989) *Annu. Rev. Biochem.* **58**, 765

Altmann, K.H., Wojcik, J., Vasquez, M. & Scheraga, H.A. (1990) *Biopolymers* **30**, 107

Anfinsen, C.B., Haber, E., Sela, M. & White, F.H. (1961) *Proc. Natn. Acad. Sci. U.S.A.* **47**, 1309

Anfinsen, C. B. (1973) *Science* **181**, 223

Baldwin, R. L. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8069

Barskaya, T.V. & Ptitsyn, O.B. (1971) *Biopolymers* **10**, 2181

Bell, J.A., Becktel, W.J., Baase, W. & Matthews, B.W. (1992) *Biochemistry* **31**, 3590

Bierzynski, A., Kim, P.S. & Baldwin, R.L. (1982) *Proc. Natn. Acad. Sci. U.S.A.* **79**, 2470

Blagdon, D.E. & Goodman, M. (1975) *Biopolymers* **14**, 241

Boisen, S., Anderson, C.Y. & Hejgaard, J. (1981) *Physiol. Plant* **52**, 167

Broadhurst, R.W., Dobson, C.M., Hore, P.J., Radford, S.E. & Rees, M.L. (1991) *Biochemistry* **30**, 405

Brown, J.E. & Klee, W.A. (1971) *Biochemistry* **10**, 470

Bruch, M. D., Dhingra, M. M. & Gierasch, L. M. (1991) *Protein Struct. Funct. Genet.* **10**, 130

Buckle, A.M., Henrick, K. & Fersht, A.R. (1993) *J. Mol. Biol.* (submitted)

Bychkova, V.E., Ptitsyn, O.B. & Barskaya, T.V. (1971) *Biopolymers* **10**, 2161

Bycroft, M., Matouschek, A., Kellis, J.T.Jr., Serrano, L. & Fersht, A.R. (1990) *Nature* **346**, 488

Campbell, A.F. (1988) *PhD Thesis*, University of London, UK.

Chakrabartty, A., Schellman, J. & Baldwin, R. (1991) *Nature* **351**, 586

Chan, H.S. & Dill, K.A. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87,** 6388

Chothia, C. (1976) *J. Mol. Biol.* **105**, 1

Chou, P.Y. & Fasman, G.D. (1974a) *Biochemistry* **13**, 211

Chou, P.Y. & Fasman, G.D. (1974b) *Biochemistry* **13**, 222

Clackson, T., Detlef, G., Jones, P.T., McPherson, M.J., ed., Quirke, P., ed., & Taylor, G.R., ed. (1991) in *PCR: A Practical Approach*, Oxford IRC Press (Oxford)

Clore, B.M., Gronenborn, A.M, Kjaer, M & Poulsen, F.M. (1987a) *Protein Engineering* **1**, 305

Clore, B.M., Gronenborn, A.M, James, M.N.G., Kjaer, M, McPhalen, C.A. & Poulsen, F.M. (1987b) *Protein Enginering* **1**, 313

Cooper, A. & Johnson, C., Jones, C., ed., Mulloy, B., ed., & Thomas, A.H., ed. (1992) in *Biological Microcalorimetry: Introductory Notes & Prcatical Procedures; Methods in Molecular Biology: Physical Methods of Analysis*, Humana Press (Clifton, N.J.)

Creighton, T.E. (1978) *Prog. Biophys. Mol. Biol.* **33**, 231

Creighton, T.E. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5082

Creighton, T.E. (1990) *Biochem. J.* **270**, 1

Creighton, T.E. (1991) *Current Biology* **1**, 8

Creighton, T.E. (1993) Proteins, 2nd ed. (W.H. Freeman & Co. New York)

Dill, K.A. (1990a) *Biochemistry* **29**, 7133

Dill, K. A. (1990b) *Science* **250**, 297

Dobson, C.M. & Evans, P.A. (1988) *Nature* **335**, 666

Eisenberg, D. & McLachlan, A. D. (1986) *Nature* **319**, 199

Ellis, R.J. & Hemmingsen, S.M. (1989) *Trends Biochem. Sci.* **14**, 339

Eriksson, A.E., Baase, W., Wozniak, J.A. & Matthews, B.W. (1992a) *Nature* **355**, 371

Eriksson, A.E., Baase, W., Zhang, X.J., Heinz, D.W., Blaber, M., Baldwin, E.P. & Matthews, B.W. (1992b) *Science* **255**, 178

Fairman, R., Armstrong, K.M., Shoemaker, K.R., York, E.J., Stewart, J.M. & Baldwin, R.L. (1991) *J.Mol. Biol.* **221**, 1395

Fauchere, J. L. & Pliska, V. E. (1983) *Eur. J. Med. Chem.* **18**, 199

Fersht, A. R. (1985) *Enzyme Structure and Mechanism*, W.H. Freeman & Co.; 2nd ed. (New York)

Fersht, A.R., Kellis, J.T. Jr., Matouschek, A. & Serrano, L. (1990) *Nature* **343**, 601

Fersht, A.R., Matouschek, A. & Serrano, L. (1992) *J. Mol. Biol.* **224**, 771

Fersht, A. R. & Serrano, L. (1993) *Current Opinion in Structural Biology* **3**, 75

Forood, B., Feliciano, E.J. & Nambiar, K.P. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 838

Freedman, R.B. (1984) *Trends Biochem. Sci.* **9**, 438

Freedman, R.B. (1989) *Cell* **57**, 1069

Gans, P.J., Pigchiang, C.L., Manning, M.C., Woody, R.W. & Kallenback, N.R. (1991) *Biopolymers* **31**, 1605

Ganter, C. & Plückthun, A. (1990) *Biochemistry* **29**, 9395

Goulobinoff, P., Gatenby, A.A. & Lorimar, G.H. (1989) *Nature* **337**, 44

Harpaz, Y., elMasry, N.F., Fersht, A.R. & Henrick, K. (1993) manuscript in progress

Harrison, S.C. & Durbin, R. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4028

Hermann, R.B. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4144

Hol, W.G.J., Van Duijnen, P.T. & Berendsen, H.J.C. (1978) *Nature* **273**, 443

Hol, W.G.J. (1985) *Progr. Biophys. Mol. Biol.* **45**, 149

Horovitz, A., Serrano, L., Avron, B., Bycrofy, M. & Fersht, A. R. (1990) *J. Mol. Biol.* **216**, 1031

Horovitz, A. & Fersht, A.R. (1990) *J. Mol. Biol.* **214**, 613

Horovitz, A. & Fersht, A.R. (1992) *J. Mol. Biol.* **224**, 733

Horovitz, A., Matthews, J. & Fersht, A.R. (1992) *J. Mol. Biol.* **227**, 560

Ikemura, H., Takagi, H. & Inouye, M. (1987) *J. Biol. Chem.* **262**, 7859

Jackson, S.E. & Fersht, A.R. (1991a) *Biochemistry* **30**, 10428

Jackson, S.E. & Fersht, A.R. (1991b) *Biochemistry* **30**, 10436

Jackson, S.E., Moracci, M., elMasry, N.F., Johnson, C. & Fersht, A.R. (1993a) *Biochemistry* (submitted)

Jackson, S.E., elMasry, N.F. & Fersht, A.R. (1993b) *Biochemistry* (submitted)

Jandu, S.K., Ray, S., Brooks, L. & Leatherbarrow, R.J. (1990) *Biochemistry* **29**, 6264

Jonassen, I. (1980) *Carlsberg Res. Commun.* **45**, 47

Kabsch, W. & Sander, C. (1983) Biopolymers **22**, 2577

Karplus, M. & Weaver, D.C. (1976) *Nature* **333**, 784

Kauzmann, W., McElroy, W.D., ed. & Glass, B., ed. (1954) *The Mechanism of Enzyme Action* , John Hopkins Press (Baltimore, MD)

Kauzmann, W. (1959) *Adv. Protein Chem.* **14**, 1

Kauzmann, W. (1987) *Nature* **325**, 763

Kellis, J.T. Jr., Nyberg, K., Sali, D. & Fersht, A.R. (1988) *Nature* **333**, 784

Kellis, J.T. Jr., Nyberg, K. & Fersht, A.R. (1989) *Biochemistry* **28**, 4914

Kemp, D.S., Boyd, J.G. & Muendel, C. C. (1991) *Nature* **352**, 451

Kiefhaber, T., Grunert, H. P., Hahn, U. & Schmid, F.X. (1990) *Biochemistry* **29**, 6475

Kim, P.S. & Baldwin, R.L. (1982) *Ann. Rev. Biochem.* **51**, 459

Kim, P.S. & Baldwin, R.L. (1984) *Nature* **307**, 329

Kitamura, S. & Sturtevant, J.M. (1989) *Biochemistry* **28**, 3788

Kjaer, M., Ludvigsen, S., Sorensen, O.W., Denys, L.A., Kindtler, J. & Poulsen, F.M. (1987) *Carlsberg Res. Commun.* **52**, 327

Kjaer, M. & Poulsen, F.M. (1987) *Carlsberg Res. Commun.* **52**, 355

Kuwajima, K., Hiroaka, Y., Ikeguchi, M. & Sugai, S. (1985) *Biochemistry* **24**, 874

Kuwajima, K. (1989) *PROTEINS Structure, Function and Genetics* **6**, 87

Laemmli, U.K. (1970) Nature **227**, 680

Lang, K., Schmid, F.X. & Fischer, G. (1987) *Nature* **329**, 268

Laskowski, M. Jr. & Kato, I. (1980) *Ann. Rev. Biochem.* **49**, 593

Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55**, 379

Lesk, A.M. & Rose, G.D. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4304

Levinthal, C.J. (1968) *J. Chem. Phy.* **65,** 44

Lifson, S., & Roig, A., (1961) *J. Chem. Phys.* **34,** 1963

Linderstrom-Lang, K.U. (1952) *Lane Medical Lectures*, **Vol. 6** pg. 53, Stanford University Press, Stanford, CA.

Longstaff, C., Campbell, A.F. & Fersht, A.R. (1990) *Biochemistry* **29**, 7339

Ludvigsen, S., Shen, H., Kjaer, M., Madsen, J. C., Poulsen, F. M., (1991) *J. Mol. Biol.* **222**, 621

Lyu, P.C., Liff, M.I., Marky, L.A. & Kallenbach, N.R. (1990) *Science* **250**, 669

Lyu, P.C., Wang, P.C., Liff, M.I. & Kallenbach, N.R. (1991) *J. Amer. Chem. Soc.* **113**, 3568

Lyu, P.C., Zhou, H., Wemmer, D.E. & Kallenbach, N.R. (1992) *J. Amer. Chem. Soc.* **114**, 6560

Lyu, P.C., Wemmer, D.E., Hongxing, X.Z., Pinker, R. & Kallenbach, N.R. (1993) *Biochemistry* **32**, 421

Marqusee, S. & Baldwin, R.L. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8898

Marqusee, S., Robbins, V.H. & Baldwin, R.L. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5286

Matouschek, A., Kellis, J.T.Jr., Serrano, L. & Fersht, A.R. (1989) *Nature* **340**, 122

Matouschek, A., Kellis, J.T. Jr., Serrano, L., Bycroft, M. & Fersht, A.R. (1990) *Nature* **346**, 440

Matouschek, A. & Fersht, A.R. (1991) *Methods Enzymol.* **202**, 82

Matouschek, A., Serrano, L. & Fersht, A.R. (1992a) *J. Mol. Biol.* **224**, 819

Matouschek, A., Serrano, L., Meiering, E.M., Bycroft, M. & Fersht, A.R. (1992b) *J. Mol. Biol.* **224**, 837

Matouschek, A. & Fersht, A.R. (1993) *Proc. Natl. Acad. Sci. (submitted)*

Matsumura, M., Becktel, W. J. & Matthews, B. W. (1988) *Nature* **334**, 406

Matthews, B. W. (1987) *Biochemistry* **26**, 6885

Matthews, B.W., Nicholson, H. & Becktel, W.J. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6663

McPhalen, C.A., Svedsen, I., Jonassen, I. & James, M.N.G. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7247

McPhalen, C.A. & James, M.N.G. (1987) *Biochemistry* **26**, 261

McPhalen, C.A. & James, M.N.G. (1988) *Biochemistry* **27**, 6582

Merutka, G. & Stellwagen, E. (1990) *Biochemistry* **29**, 894

Merutka, G., Lipton, W., Shalongo, W., Park, S.H. & Stellwagen, E. (1990) *Biochemistry* **29**, 7511

Miller, S., Janin, J., Lesk, A. M. & Chothia, C. (1987) *J. Mol. Biol.* **196**, 641

Mitchison, C. & Baldwin, R.L. (1986) *Proteins* **1**, 23

Moult, J. & Unger, R. (1991) *Biochemistry* **30**, 3816

Mullis, K.B. & Faloon, F.A. (1987) *Methods Enzymol.* **155**, 335

Munck, L., Karlson, K.E., Hagberg, A. & Eggum, B.O. (1970) *Science* **168**, 985

Nakai, K., Kidera, A. & Kanehisa, M. (1988) *Protein Eng.* **2**, 93

Nicholson, H., Becktel, W.J. & Matthews, B.W. (1988) *Nature* **336**, 651

Nicholson, H., Anderson, D.E., Dao-pin, S. & Matthews, B.W. (1991) *Biochemistry* **30**, 9816

Oas, T.G. & Kim, P.S. (1988) *Nature* **336**, 42

Ooi, T., Oobatake, M., Nemethy, G. & Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3086

Ooi, T., Oobatake, M. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 2859

O'Neil, K.T. & DeGrado, W.F. (1990) *Science* **250**, 646

Pace, C.N. (1986) *Methods Enzymol.* **131**, 266

Pace, C.N. (1992) *J. Mol. Biol.* **226**, 29

Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue & Baldwin, R.L. (1990) *Nature* **344**, 268

Padmanabhan, S. & Baldwin, R.L. ((1991) *J.Mol. Biol.* **219**, 135

Presta, L.G. & Rose, G.D. (1988) *Science* **240**, 1632

Privalov, P.L. (1979) *Adv. Protein Chem.* **33**, 167

Privalov, P.L. (1986) *Methods Enzymol.* **131**, 4

Privalov, P. L. & Gill, S. J. (1988) *Adv. Protein Chem.* **39**, 191

Ptitsyn, O.B. & Finklestein, A.V. (1983) *Biopolymers* **22**, 15

Ptitsyn, O.B. (1987) *J. Protein Chem.* **6**, 273

Ptitsyn, O.B., Pain, R.H., Semisotnov, G.V., Zerovnik, E. & Razulyaev, O.I. (1990) *FEBS Lett.* **12**, 20

Radzicka, A. & Wolfenden, R. (1988) *Biochemistry* **27**, 1664

Ramachandran, G. N. & Sasisekharan, V. (1968) *Adv. Protein Chem.* **23**, 283

Richardson, J.S. & Richardson, D.C. (1988) *Science* **240**, 1648

Robson, B. & Garnier, J. (1988) *Introduction to Proteins and Protein Engineering*, Elsevier Press (Netherlands)

Roder, H., Elöve, G.A. & Englander, S.W. (1988) *Nature* **335**, 694

Rose, G.D., Geselowitz, A.R., Lesser, G.J., Lee, R.H. & Zehfus, M. (1985) *Science* **229**, 834

Ryan, C.A. (1978) *Trends Biochem. Sci.* **3**, 148

Sali, D., Bycroft, M. & Fersht, A.R. (1988) *Nature* **335**, 740

Sambrook, J., Fritsh, E.F. & Maniatis, T., eds. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Labortatory Press; 2nd. ed. (New York)

Sancho, J. & Fersht, A.R. (1992a) *J. Mol. Biol.* **224**, 741

Sancho, J., Neira, J.L. & Fersht, A.R. (1992b) *J. Mol. Biol.* **224**, 749

Sandberg, W. S. & Terwilliger, T. C. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1706

Sanger, F., Nicklen, S. & Coulson, A.R. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463

Sayers, J.R., Eckstein, F. & Setlow, J.K., ed. (1988) *Genetic Engineering: Principles and Methods* **10**, pg. 109 , Plenum Press (New York and London)

Schmid, F.X., Creighton, T.E., ed. (1989) in *Protein Structure: A Practical Approach* I.R.L. Press at Oxford University Press (Oxford)

Schmid, F.X. (1991) *Current Opinion in Structural Biology* **1**, 36

Schulz, G. E. & Shirmer, R. H. (1979) *Principles in Protein Structure*, Springer (New York)

Seemuller, U., Meier, M., Ohlsson, K., Muller, H.P. & Fritz, H. (1977) *Hoppe-Seylers Z-Physiol. Chem.* **358**, 1105

Serrano, L. & Fersht, A.R. (1989) *Nature* **342**, 296

Serrano, L., Kellis, J.T. Jr., Cann, P., Matouschek, A. & Fersht, A.R. (1992a) *J. Mol. Biol.* **224**, 783

Serrano, L., Matouschek, A. & Fersht, A.R. (1992b) *J. Mol. Biol.* **224**, 805

Serrano, L., Sancho, J., Hirshberg, M. & Fersht, A.R. (1992c) *J. Mol. Biol.* **227**, 544

Serrano, L., Niera, J.L., Sancho, J. & Fersht, A.R. (1992d) *Nature* **356**, 453

Sharp, K. A., Nicholls, A., Friedman, R. & Honig, B. (1991) *Biochemistry* **30**, 9686

Sheridan, R.P., Levy, R.M. & Salemme, F.R. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4545

Shoemaker, K.R., Kim, P.S., York, E.J., Stewart, J.M. & Baldwin, R.L. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2349

Shoemaker, K.R., Kim, P.S., York, E.J., Stewart, J.M. & Baldwin, R.L. (1987) *Nature* **326**, 563

Shortle, D. & Meeker, A.K. (1989) *Biochemistry* **28**, 936

Shortle, D., Stites, W. E. & Meeker, A. K. (1990) *Biochemistry* **29**, 8033

Shrake, A. & Ross, P.D. (1990) *Nature* **341**, 462

Staley, J.P. & Kim, P.S. (1990) *Nature* **344**, 685

Strehlow, K.G & Baldwin, R.L. (1989) *Biochemistry* **28**, 2130

Strehlow, K.G., Robertson, A.D. & Baldwin, R.L. (1991) *Biochemistry* **30**, 5801

Sueki, M., Lee, S., Powers, S.P., Denton, J.B., Konishi, Y. & Scheraga, H.A. (1984) *Macromolecules* **17**, 148

Svendsen, I., Jonassen, I., Hejgaard, J., & Boisen, S. (1980) *Carlsberg Res. Commun.* **45**, 389

Svendsen, I., Boisen, S. & Hejgaard, J. (1982) *Carlsberg Res. Commun.* **47**, 45

Tanford, C.H. (1962) *J. Phys. Chem.* **84**, 4240

Tanford, C.H. (1968) *Adv. Protein Chem.* **23**, 121

Tanford, C.H. (1970) *Adv. Protein Chem.* **24**, 1

Udgoankar, J.B. & Baldwin, R.L. (1988) *Nature* **335**, 700

Viera,J. & Messing, J. (1987) *Methods Enzymol.* **153**, 3

Wada, A. (1976) *Adv. Biophys.* **9**, 1

Winder, A.F. & Gent, W.L.G. (1971) *Biopolymers* **10**, 1243

Yutani, K., Ogasahara, K., Tsujita, T. & Sugino, Y. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4441

Zhu, X., Ohta, Y., Jordan, F. & Inouye, M. (1989) *Nature* **339**, 483

Zimm, B. H., & Bragg, J. R. (1959) *J. Chem. Phys.* **31,** 526

Zwanzig, R., Szabo, A. & Bagchi, B. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 20

Appendix I

Statistics, Errors of Observation and Accuracy§

A.1 Normal or Gaussian distribution

The distribution of errors of measurement are usually analysed according the Gaussian or normal distribution. This applies to sampling a population that is subject to a random distribution. The normal distribution follows the equation:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\bar{x})^2/2\sigma^2} \qquad (1)$$

The quantity σ is called the *standard deviation* of the population and is a measure of the spread or dispersion of the values about the mean (see Table). The smaller is σ, the narrower the spread. σ is calculated from: the equation:

$$\sigma = \left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right]^{1/2} \qquad (2)$$

TABLE **Probability of x being within various limits of its mean value ($\bar{x}$)**

Range of values	Probability of x within this range
$\bar{x} \pm 0.6745\ \sigma$	0.5
$\bar{x} \pm \sigma$	0.683
$\bar{x} \pm 2\sigma$	0.954
$\bar{x} \pm 3\sigma$	0.997

A.2 Errors in sampling

Suppose one takes n measurements of a quantity that are subject to random errors. The value of the mean so measured follows a normal distribution about the true mean. Most calculators give a standard deviation calculated from the formula:

$$\sigma_{calc} = \left[\frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2 \right]^{1/2} \quad (3)$$

(which is eqn 2 modified slightly). However, the important quantity for the calculation of accuracy is the *standard error of the mean*, α, where:

$$\alpha = \sigma_{calc}/n^{1/2} \quad (4)$$

The relationship between distribution of the measured value of $\bar{x}$ and α follows exactly the same distribution as in the table. Thus, the probability of the true mean being between $\bar{x} \pm \alpha$ is 0.683, between $\bar{x} \pm 2\alpha$ is 0.954, and $\bar{x} \pm 3\alpha$ is 0.997. Note that the value of α has itself a standard error that depends on n: $\alpha_\alpha = \alpha/(2n)^{1/2}$.

A.3 Combining errors of measurement

a. Sum or difference.

Suppose we have a quantity y which is the sum or difference of two other quantities, a, and b, and each has a standard error of α_a or α_b, then, the standard error of y is given by:

$$\alpha_y = (\alpha_a^2 + \alpha_b^2)^{1/2/} \quad (5)$$

This may be extended to any series of a, b, c etc:

$$\alpha_y = [(\alpha_a^2 + \alpha_b^2 + \alpha_c^2 .....)^{1/2}]/n \quad (6)$$

b. Product or quotient.

Suppose $y = ab$, then:

$$\alpha_y = (b^2\alpha_a^2 + a^2\alpha_b^2)^{1/2} \quad (7)$$

For more complex multiplications it is easier to calculate the error as a fraction. For example, if $y = abc$, then:

$$\alpha_y/y = [(\alpha_a/a)^2 + (\alpha_b/b)^2 + (\alpha_c^2/c)]^{1/2} \quad (8)$$

Equation 8 may be manipulated for a power law, $y = a^n$, to give:

$$\alpha_y/y = n\alpha_a/a \quad (9)$$

or for a quotient, $y = a/b$, to give:

$$\alpha_y/y = [(\alpha_a/a)^2 + (\alpha_b b)^2]^{1/2} \quad (10)$$

c. logarithm.

From elementary calculus, small changes, δ, in a natural logarithm of a quantity y and in y are related by: $\delta \ln y = \delta y/y$.

Consequently,

$$\alpha_{\ln y} = \alpha_y / y \tag{11}$$

Since log (to base 10) and ln are related by: $\ln y = 2.303 \log y$,

$$\alpha_{\log y} = (\alpha_y / y)/2.303 \tag{12}$$

d. Weighting data.

Measurements that have a low standard error are clearly more significant than those with a large standard error. Each measurements should be weighted proportionally to the reciprocal of the square of its standard error. Suppose there n measurements of a value and each value x_i has a standard error α_i, then the weighted mean value is given by:

$$\bar{x} = \left(\sum x_i / \alpha_i^2\right) / \left(\sum 1 / \alpha_i^2\right) \tag{13}$$

The standard error of the weighted mean is given by:

$$\alpha = \left[\frac{1}{(n-1)\sum (1/\alpha_i^2)} \sum \left(x_i - \bar{x}\right)^2 / \alpha_i^2\right]^{1/2} \tag{14}$$

A.4 Poisson distribution.

The statistics of processes such as radioactive decay and emission of light that produce a flux of particles or *distributive* polymerases that add residues at random to growing polymer chains obey the Poisson distribution. The number of particles measured per unit time or the number of residues added to a particular chain varies about the mean value $\bar{x}$ according to eqn 15.

$$p(k) = \frac{\bar{x}^k}{k!} e^{\bar{x}} \tag{15}$$

where $p(k)$ is the probability of k particles being emitted or residues added. The Gaussian and Poisson distributions approximate to each other when there is a large number of particles involved in the Poisson.

The standard deviation of the Poisson distribution is equal to $\bar{x}^{1/2}$. This has profound consequences in designing experiments involving radioactivity and light intensity.

A.5 Signal to noise in Absorbance, Circular Dichroism, Fluorescence, and Radioactive Counting

There is thus an inherent noise in measuring a radioactive flux or light intensity, say n, that is proportional to $\sqrt{n}$. in because there are fluctuations in its apparent strength. Suppose the noise, N, = $\beta\sqrt{n}$. Then the ratio of noise to signal, (N/S) is given by:

$$(N/S) = \beta/n^{1/2} \tag{16}$$

Thus the greater is n, the lower is the (N/S) ratio. For processes such as radioactive decay and fluorescence, the greater the specific activity or the the greater the intensity of the exciting light, the lower the is the error or noise in the measurements. Similarly, the longer the period of counting, the more the precision. Rapid reaction experiments where optical signals are collected over shorter times are inherently noisy because the shorter the time range, the smaller the number of particles counted. Accuracy improves according to the square root of the time of collection or the time constant of the electronic filter that is used to smooth the signal. Absorbance and CD measurements respond in a more complex way because of the logarithmic relationship of absorbance to the intensity of transmitted light: there is a balance between the advantage of having a high absorbance and the loss of signal. According to Beer's law, $A = \log (I_o/I)$, where A is the absorbance, I_0 the light intensity entering the solution and I the transmitted light. Thus:

$$I = I_0 10^{-A} \tag{17}$$

From eqn 16, the ratio of the noise in transmitted intensity, δI, to I is given by:

$$\delta I/I = \pm(\beta 10^{A/2})/I_0^{1/2} \tag{18}$$

The relationship between fluctuations in A (δA) and those in I may be derived as for eqn 12 to be $\delta A = \delta I/2.303I$. The ratio of signal to noise in A, $(N/S)_A$, is given by $\delta A/A$. Thus,

$$(N/S)_A = \pm\delta A/A = \pm(\beta 10^{A/2})/2.303A\sqrt{I_0} \tag{19}$$

The minimum is found by differentiating $(N/S)_A$ with respect to A. The absorbance where the value of $(N/S)_A$ is at a minimum is given by:

$$A_{min} = 2/2.303 = 0.87 \quad (20)$$

The optimal absorbance to use in absorbance or CD measurements for best ratio of signal to noise is 0.87.

[§] reproduced with the very kind permission of Prof. A.R. Fersht, MRC Unit of Protein Function and Design.